A Shooting Algorithm for Optimal Control Problems with Singular Arcs

Maria Soledad Aronna, J. Frederic Bonnans, Pierre Martinon

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A shooting algorithm for problems with singular arcs

M. Soledad Aronna — J. Frédéric Bonnans — Pierre Martinon

N° 7763 — version 2

initial version Octobre 2011 — revised version June 2012

Thème NUM

ISSN 0249-6399 ISRN INRIA/RR--7763--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
SACLAY - ÎLE-DE-FRANCE

A shooting algorithm for problems with singular arcs

M. Soledad Aronna*†, J. Frédéric Bonnans †, Pierre Martinon †

Thème NUM — Systèmes numériques
Équipes-Projets Commands

Rapport de recherche n° 7763 — version 2 — initial version Octobre 2011 — revised version June 2012 — 40 pages

Abstract: In this article we propose a shooting algorithm for a class of optimal control problems for which all control variables appear linearly. The shooting system has, in the general case, more equations than unknowns and the Gauss-Newton method is used to compute a zero of the shooting function. This shooting algorithm is locally quadratically convergent if the derivative of the shooting function is one-to-one at the solution. The main result of this paper is to show that the latter holds whenever a sufficient condition for weak optimality is satisfied. We note that this condition is very close to a second order necessary condition. For the case when the shooting system can be reduced to one having the same number of unknowns and equations (square system) we prove that the mentioned sufficient condition guarantees the stability of the optimal solution under small perturbations and the invertibility of the Jacobian matrix of the shooting function associated to the perturbed problem. We present numerical tests that validate our method.

Key-words: optimal control, Pontryagin Maximum Principle, singular control, constrained control, shooting algorithm, second order optimality condition, stability

This work is supported by the European Union under the 7th Framework Programme FP7-PEOPLE-2010-ITN Grant agreement number 264735-SADCO

* CIFASIS-CONICET Argentina (aronna@cmap.polytechnique.fr)
† INRIA-Saclay and CMAP, École Polytechnique, 91128 Palaiseau, France (Frederic.Bonnans@inria.fr) (Pierre.Martinon@inria.fr)

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université
4, rue Jacques Monod, 91893 ORSAY Cedex
Téléphone : +33 1 72 92 59 00

Un algorithme de tir pour les problèmes de commande optimale avec des arcs singuliers

Résumé : Dans ce travail on présente une condition suffisante pour que l'algorithme de tir soit localement convergent quand il est appliqué aux problèmes de commande optimale affines dans les commandes. On commence par étudier le cas avec des contraintes initiales-finales sur l'état et commande libre, et en suite on ajoute des contraintes sur la commande. L'algorithme de tir est localement quadratiquement convergent si la dérivée de la fonction de tir associée est injective dans la solution optimale. Le résultat principal de cet article montre une condition suffisante pour cette injectivité, qui est très proche de la condition nécessaire du second ordre. On montre que cette condition suffisante assure la stabilité de la solution optimale aux petites perturbations et qu'elle garantit aussi que l'algorithme de tir est convergent pour le problème perturbé. On présente des essais numériques qui valident notre méthode.

Mots-clés : commande optimale, Principe de Pontryaguine, commande singulière, contraintes sur la commande, algorithme de tir, conditions d'optimalité du second ordre, stabilité

1 Introduction

The classical shooting method is used to solve boundary value problems. Hence, it is used to compute the solution of optimal control problems by solving the boundary value problem derived from the Pontryagin Maximum Principle.

Some references can be mentioned regarding the shooting method. The first two works we can find in the literature, dating from years 1956 and 1962 respectively, are Goodman-Lance [1] and Morrison et al. [2]. Both present the same method for solving two-point boundary value problems in a general setting, not necessarily related to an optimal control problem. The latter article applies to more general formulations. The method was studied in detail in Keller's book [3], and later on Bulirsch [4] applied it to the resolution of optimal control problems.

The case we deal with in this paper where the shooting method is used to solve optimal control problems with control-affine systems is treated in, e.g., Maurer [5], Oberle [6, 7], Fraser-Andrews [8], Martinon [9] and Vossen [10]. These works provided a series of algorithms and numerical examples with different control structures, but no theoretical foundation is supplied. In particular, Vossen [10] dealt with a problem in which the control can be written as a function of the state variable, i.e. the control has a feedback representation. He proposed an algorithm that involved a finite dimensional optimization problem induced by the switching times. The main difference between Vossen's work and the study here presented is that we treat the general problem (no feedback law is necessary). Furthermore we justify the well-definition and the convergence of our algorithm via second order sufficient conditions of the original control problem. In some of the just mentioned papers the control variable had only some of its components entering linearly. This particular structure is studied in more detailed in Aronna [11], and in the present article we study problems having all affine inputs.

In [12] Bonnard and Kupka studied the optimal time problem of a generic single-input affine system without control constraints, with fixed initial point and terminal point constrained to a given manifold. For this class of problems they established a link between the injectivity of the shooting function and the optimality of the trajectory by means of the conjugate and focal points theory. Bonnard et al. [13] provides a survey on a series of algorithms for the numerical computation of these points, which can be employed to test the injectivity of the shooting function in some cases. The reader is referred to [13], Bonnard-Chyba [14] and references therein for further information about this topic.

In addition, Malanowski-Maurer [15] and Bonnans-Hermant [16] dealt with a problem having mixed control-state and pure state running constraints and satisfying the strong Legendre-Clebsch condition (which is not verified in our affine-input case). They all established a link between the invertibility of the Jacobian of the shooting function and some second order sufficient condition for optimality. They provided stability analysis as well.

We start this article by presenting an optimal control problem affine in the control, with terminal constraints and free control variables. For this kind of problem we state a set of optimality conditions which is equivalent to the Pontryagin Maximum Principle. Afterwards, the second order strengthened generalized Legendre-Clebsch condition is used to eliminate the control variable from the stationarity condition. The resulting set of conditions turns out to be a two-

point boundary value problem, i.e. a system of ordinary differential equations having boundary conditions both in the initial and final times. We define the shooting function as the mapping that assigns to each estimate of the initial values, the value of the final condition of the corresponding solution. The shooting algorithm consists of approximating a zero of this function. In other words, the method finds suitable initial values for which the corresponding solution of the differential equation system satisfies the final conditions.

Since the number of equations happens to be, in general, greater than the number of unknowns, the Gauss-Newton method is a suitable approach for solving this overdetermined system of equations. The reader is referred to Dennis [17], Fletcher [18] and Dennis et al. [19] for details and implementations of Gauss-Newton technique. This method is applicable when the derivative of the shooting function is one-to-one at the solution, and in this case it converges locally quadratically.

The main result of this paper is to provide a sufficient condition for the injectivity of this derivative, and to notice that this condition is quite weak since, for qualified problems, it characterizes quadratic growth in the weak sense (see Dmitruk [20, 21]). Once the unconstrained case is investigated, we pass to a problem having bounded controls. To treat this case, we perform a transformation yielding a new problem without bounds, we prove that an optimal solution of the original problem is also optimal for the transformed one and we apply our above-mentioned result to this modified formulation.

It is interesting to mention that by means of the latter result we can justify, in particular, the invertibility of the Jacobian of the shooting function proposed by Maurer [5]. In this work, Maurer suggested a method to treat problems having scalar bang-singular-bang solutions and provided a square system of equations (i.e. a system having as many equations as unknowns) meant to be solved by Newton's algorithm. However, the systems that can be encountered in practice may not be square and hence our approach is suitable.

We provide a deeper analysis in the case when the shooting system can be reduced to one having equal number of equations and unknowns. In this framework, we investigate the stability of the optimal solution. It is shown that the above-mentioned sufficient condition guarantees the stability of the optimal solution under small perturbation of the data and the invertibility of the Jacobian of the shooting function associated to the perturbed problem. Felgenhauer in [22, 23] provided sufficient conditions for the stability of the structure of the optimal control, but assuming that the perturbed problem had an optimal solution.

Our article is organized as follows. In section 2 we present the optimal control problem without bound constraints, for which we provide an optimality system in section 3. We give a description of the shooting method in section 4. In section 5 we present a set of second order necessary and sufficient conditions, and the statement of the main result. We introduce a linear quadratic optimal control problem in section 6. In section 7 we present a variable transformation relating the shooting system and the optimality system of the linear quadratic problem mentioned above. In section 8 we deal with the control constrained case. A stability analysis for both unconstrained and constrained control cases is provided in section 9. Finally we present some numerical tests in section 10, and we devote section 11 to the conclusions of the article.

2 Statement of the Problem

Consider the spaces $\mathcal{U} := L_\infty(0,T;\mathbb{R}^m)$ and $\mathcal{X} := W^1_\infty(0,T;\mathbb{R}^n)$, as control and state spaces, respectively. Denote by u and x their elements, respectively. When needed, put $w = (x,u)$ for a point in the product space $\mathcal{W} := \mathcal{X} \times \mathcal{U}$. In this paper we investigate the optimal control problem

$$J := \varphi_0(x_0, x_T) \to \min, \tag{1}$$

$$\dot{x}_t = \sum_{i=0}^{m} u_{i,t} f_i(x_t), \quad \text{a.e. on } [0,T], \tag{2}$$

$$\eta_j(x_0, x_T) = 0, \quad \text{for } j = 1, \dots, d_\eta, \tag{3}$$

where final time T is fixed, $u_0 \equiv 1$, $f_i : \mathbb{R}^n \to \mathbb{R}^n$ for $i = 0, \dots, m$ and $\eta_j : \mathbb{R}^{2n} \to \mathbb{R}$ for $j = 1, \dots, d_\eta$. Assume that data functions φ_0, f_i and η_j have Lipschitz-continuous second derivatives. Denote by (P) the problem defined by (1)-(3). An element $w \in \mathcal{W}$ satisfying (2)-(3) is called a *feasible trajectory*.

Set $\mathcal{X}_* := W^1_\infty(0,T;\mathbb{R}^{n,*})$ the space of Lipschitz-continuous functions with values in the $n-$dimensional space of row-vectors with real components $\mathbb{R}^{n,*}$. Consider an element $\lambda := (\beta, p) \in \mathbb{R}^{d_\eta,*} \times \mathcal{X}_*$ and define the *pre-Hamiltonian* function

$$H[\lambda](x,u,t) := p_t \sum_{i=0}^{m} u_i f_i(x), \tag{4}$$

the *initial-final Lagrangian* function

$$\ell[\lambda](\zeta_0, \zeta_T) := \varphi_0(\zeta_0, \zeta_T) + \sum_{j=1}^{d_\eta} \beta_j \eta_j(\zeta_0, \zeta_T), \tag{5}$$

and the *Lagrangian* function

$$\mathcal{L}[\lambda](w) := \ell[\lambda](x_0, x_T) + \int_0^T p_t \left(\sum_{i=0}^{m} u_{i,t} f_i(x_t) - \dot{x}_t \right) \mathrm{d}t. \tag{6}$$

We study a nominal feasible trajectory $\hat{w} = (\hat{x}, \hat{u})$. Next we present a *qualification hypothesis* that is assumed throughout the article. Consider the mapping

$$\begin{array}{rcl} G\colon \mathbb{R}^n \times \mathcal{U} & \to & \mathbb{R}^{d_\eta} \\ (x_0, u) & \mapsto & \eta(x_0, x_T), \end{array} \tag{7}$$

where x_T is the solution of (2) associated to (x_0, u).

Assumption 2.1. *The derivative of G at $(\hat{x}_0, \hat{u})$ is onto.*

Assumption 2.1 is usually known as *qualification of equality constraints.*

Definition 2.2. *It is said that the trajectory $\hat{w}$ is a weak minimum of problem (P) if there exists $\varepsilon > 0$ such that $\hat{w}$ is a minimum in the set of feasible trajectories $w = (x,u) \in \mathcal{W}$ satisfying*

$$\|x - \hat{x}\|_\infty < \varepsilon, \quad \|u - \hat{u}\|_\infty < \varepsilon.$$

The following first order necessary condition holds for $\hat{w}$.

Theorem 2.3. *If $\hat{w}$ is a weak solution then there exists $\lambda = (\beta, p) \in \mathbb{R}^{d_\eta,*} \times \mathcal{X}_*$ such that p is solution of the costate equation*

$$-\dot{p}_t = D_x H[\lambda](\hat{x}_t, \hat{u}_t, t), \quad \text{a.e. on } [0,T], \tag{8}$$

with transversality conditions

$$p_0 = -D_{x_0}\ell[\lambda](\hat{x}_0, \hat{x}_T), \tag{9}$$
$$p_T = D_{x_T}\ell[\lambda](\hat{x}_0, \hat{x}_T), \tag{10}$$

and the stationarity condition

$$D_u H[\lambda](\hat{x}_t, \hat{u}_t, t) = 0, \quad \text{a.e. on } [0,T], \tag{11}$$

is verified.

It follows easily that since the pre-Hamiltonian H is affine in all the control variables, (11) is equivalent to the *minimum condition*

$$H[\lambda](\hat{x}_t, \hat{u}_t, t) = \min_{v \in \mathbb{R}^m} H[\lambda](\hat{x}_t, v, t), \quad \text{a.e. on } [0,T]. \tag{12}$$

In order words, the element $(\hat{w}, \lambda)$ in Theorem 2.3 satisfies the *qualified Pontryagin Maximum Principle* and λ is a *Pontryagin multiplier.* It is known that the Assumption 2.1 implies the existence and uniqueness of multiplier. We denote this unique multiplier by $\hat{\lambda} = (\hat{\beta}, \hat{p})$.

Let the *switching function* $\Phi : [0,T] \to \mathbb{R}^{m,*}$ be defined by

$$\Phi_t := D_u H[\hat{\lambda}](\hat{x}_t, \hat{u}_t, t) = (\hat{p}_t f_i(\hat{x}_t))_{i=1}^m. \tag{13}$$

Observe that the stationarity condition (11) can be written as

$$\Phi_t = 0, \quad \text{a.e. on } [0,T]. \tag{14}$$

3 Optimality System

In this section we present an optimality system, i.e. a set of equations that are necessary for optimality. We obtain this system from the conditions in Theorem 2.3 above and assuming that the *strengthened generalized Legendre-Clebsch condition* (to be defined below) holds.

Observe that, since H is affine in the control, the switching function Φ introduced in (13) does not depend explicitly on u. Let an index $i = 1, \dots, m$, and $(\mathrm{d}^{M_i}\Phi/\mathrm{d}t^{M_i})$ be the lowest order derivative of Φ in which u_i appears with a coefficient that is not identically zero on $(0,T)$. In Kelley et al. [24] it is stated that M_i is even, assuming that the extremal is normal (as it is the case here since $\hat{w}$ satisfies the PMP in its qualified form). The integer $N_i := M_i/2$ is called *order of the singular arc.* As we have just said, the control u cannot be retrieved from equation (11). In order to be able to express $\hat{u}$ in terms of $(\hat{p}, \hat{x})$ from equation

$$\ddot{\Phi}_t = 0, \quad \text{a.e. on } [0,T], \tag{15}$$

we make the following hypothesis.

Assumption 3.1. *The strengthened generalized Legendre-Clebsch condition (see e.g.*
Kelley [25] and Goh [26]) holds, i.e.

$$-\frac{\partial}{\partial u}\ddot{\Phi}_t \succ 0, \quad \text{on } [0,T]. \tag{16}$$

Here, by $X \succ 0$ we mean that the matrix X is positive definite. Notice that function $\ddot{\Phi}$ is affine in u, and thus $\hat{u}$ can be written in terms of $(\hat{p},\hat{x})$ from (15) by inverting the matrix in (16). Furthermore, due to the regularity hypothesis imposed on the data functions, $\hat{u}$ turns out to be a continuous function of time.

Hence, the condition (15) is included in our optimality system and we can use it to compute $\hat{u}$ in view of Assumption 3.1. In order to guarantee the stationarity condition (14) we consider the endpoint conditions

$$\Phi_T = 0, \ \dot{\Phi}_0 = 0. \tag{17}$$

Remark 3.2. *We could choose another pair of endpoint conditions among the four possible ones:* $\Phi_0 = 0$, $\Phi_T = 0$, $\dot{\Phi}_0 = 0$ *and* $\dot{\Phi}_T = 0$, *always including at least one of order zero. The choice we made will simplify the presentation of the result afterwards.*

Notation: Denote by (OS) the set of equations composed by (2)-(3),(8)-(10), (15), (17), i.e. the system

$$\begin{cases} \dot{x}_t = \sum_{i=0}^{m} u_{i,t} f_i(x_t), \quad \text{a.e. on } [0,T], \\ \eta_j(x_0,x_T) = 0, \quad \text{for } j = 1,\dots,d_\eta, \\ -\dot{p}_t = D_x H[\lambda](\hat{x}_t,\hat{u}_t,t), \quad \text{a.e. on } [0,T], \\ p_0 = -D_{x_0}\ell[\lambda](\hat{x}_0,\hat{x}_T), \\ p_T = D_{x_T}\ell[\lambda](\hat{x}_0,\hat{x}_T), \\ \ddot{\Phi}_t = 0, \quad \text{a.e. on } [0,T], \\ \Phi_T = 0, \ \dot{\Phi}_0 = 0. \end{cases} \tag{OS}$$

Let us give explicit expressions for $\dot{\Phi}$ and $\ddot{\Phi}$. Denote the *Lie bracket* of two smooth vector fields $g, h\colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$[g,h](x) := g'(x)h(x) - h'(x)g(x). \tag{18}$$

Define $A\colon \mathbb{R}^{n+m} \to \mathcal{M}_{n\times n}(\mathbb{R})$ and $B\colon \mathbb{R}^n \to \mathcal{M}_{n\times m}(\mathbb{R})$ by

$$A(x,u) := \sum_{i=0}^{m} u_i f_i'(x), \quad B(x)v := \sum_{i=1}^{m} v_i f_i(x), \tag{19}$$

for every $v \in \mathbb{R}^m$. Notice that the ith. column of $B(x)$ is $f_i(x)$. For $(x,u) \in \mathcal{W}$ satisfying (2), let $B_1(x_t,u_t) \in \mathcal{M}_{n\times m}(\mathbb{R})$ given by

$$B_1(x_t,u_t) := A(x_t,u_t)B(x_t) - \frac{\mathrm{d}}{\mathrm{d}t}B(x_t). \tag{20}$$

In view of (19) and (20), the expressions in (17) can be rewritten as

$$\Phi_t = p_t B(x_t), \quad \dot{\Phi}_t = -p_t B_1(x_t,u_t). \tag{21}$$

4 Shooting Algorithm

The aim of this section is to present an appropriated numerical scheme to solve system (OS). For this purpose define the *shooting function*

$$\begin{aligned}\mathcal{S}\colon \mathrm{D}(\mathcal{S}) := \mathbb{R}^n \times \mathbb{R}^{n+d_\eta,*} \quad &\to \mathbb{R}^{d_\eta} \times \mathbb{R}^{2n+2m,*},\\ \big(x_0, p_0, \beta\big) =: \nu \quad &\mapsto \mathcal{S}(\nu) := \begin{pmatrix} \eta(x_0, x_T) \\ p_0 + D_{x_0}\ell[\lambda](x_0, x_T) \\ p_T - D_{x_T}\ell[\lambda](x_0, x_T) \\ p_T B(x_T) \\ p_0 B_1(x_0, u_0) \end{pmatrix},\end{aligned} \tag{22}$$

where (x, u, p) is a solution of (2),(8),(15) corresponding to the initial conditions (x_0, p_0), and with $\lambda := (\beta, p)$. Here we denote either by (a_1, a_2) or $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ an element of the product space $A_1 \times A_2$. Notice that the control u retrieved from (15) is continuous in time, as we have already pointed out after Assumption 3.1. Hence, we can refer to the value u_0, as it is done in the right hand-side of (22). Observe that in a simpler framework having fixed initial state and no final constraints, the shooting function would depend only on p_0. In our case, since the initial state is not fixed and a multiplier associated with the initial-final constraints must be considered, $\mathcal{S}$ has more independent variables. Note that solving (OS) consists of finding $\nu \in \mathrm{D}(\mathcal{S})$ such that

$$\mathcal{S}(\nu) = 0. \tag{23}$$

Since the number of equations in (23) is greater than the number of unknowns, the Gauss-Newton method is a suitable approach to solve it. This algorithm will solve the equivalent least squares problem

$$\min_{\nu \in \mathrm{D}(\mathcal{S})} \left|\mathcal{S}\left(\nu\right)\right|^2. \tag{24}$$

At each iteration k, given the approximate values ν^k, it looks for Δ^k that gives the minimum of the linear approximation of problem

$$\min_{\Delta \in \mathrm{D}(\mathcal{S})} \left|\mathcal{S}(\nu^k) + \mathcal{S}'(\nu^k)\Delta\right|^2. \tag{25}$$

Afterwards it updates

$$\nu^{k+1} \leftarrow \nu^k + \Delta^k. \tag{26}$$

In order to solve the linear approximation of problem (25) at each iteration k, we look for Δ^k in the kernel of the derivative of the objective function, i.e. Δ^k satisfying

$$\mathcal{S}'(\nu^k)^\top \mathcal{S}'(\nu^k)\Delta^k + \mathcal{S}'(\nu^k)^\top \mathcal{S}(\nu^k) = 0. \tag{27}$$

Hence, to compute direction Δ^k the matrix $\mathcal{S}'(\nu^k)^\top \mathcal{S}'(\nu^k)$ must be nonsingular. Thus, Gauss-Newton method will be applicable provided that $\mathcal{S}'(\hat{\nu})^\top \mathcal{S}'(\hat{\nu})$ is invertible, where $\hat{\nu} := (\hat{x}_0, \hat{p}_0, \hat{\beta})$. Easily follows that $\mathcal{S}'(\hat{\nu})^\top \mathcal{S}'(\hat{\nu})$ is nonsingular if and only if $\mathcal{S}'(\hat{\nu})$ is one-to-one. Summarizing, the *shooting algorithm* we propose here consists of solving the equation (23) by the Gauss-Newton method defined by (26)-(27).

Since the right hand-side of system (23) is zero, the Gauss-Newton method converges locally quadratically if the function $\mathcal{S}$ has Lipschitz-continuous derivative. The latter holds here given the regularity assumptions on the data functions. This convergence result is stated in the proposition below. See, e.g., Fletcher [18] or Bonnans [27] for a proof.

Proposition 4.1. *If $\mathcal{S}'(\hat{\nu})$ is one-to-one then the shooting algorithm is locally quadratically convergent.*

The main result of this article is to present a condition that guarantees the quadratic convergence of the shooting method near the optimal local solution $(\hat{w}, \hat{\lambda})$. This condition involves the second variation studied in Dmitruk [20, 21], more precisely, the sufficient optimality conditions therein presented.

4.1 Linearization of a Differential Algebraic System

For the aim of finding an expression of $\mathcal{S}'(\hat{\nu})$, we make use of the linearization of (OS) and thus we introduce the following concept.

Definition 4.2 (Linearization of a Differential Algebraic System)**.** *Consider a system of differential algebraic equations (DAE) with endpoint conditions*

$$\dot{\zeta}_t = \mathcal{F}(\zeta_t, \alpha_t), \tag{28}$$
$$0 = \mathcal{G}(\zeta_t, \alpha_t), \tag{29}$$
$$0 = \mathcal{I}(\zeta_0, \zeta_T), \tag{30}$$

where $\mathcal{F} : \mathbb{R}^{m+n} \to \mathbb{R}^n$, $\mathcal{G} : \mathbb{R}^{m+n} \to \mathbb{R}^{d_\mathcal{G}}$ and $\mathcal{I} : \mathbb{R}^{2n} \to \mathbb{R}^{d_\mathcal{I}}$ are C^1 functions. Let (ζ^0, α^0) be a C^1 solution. We call linearized system at point (ζ^0, α^0) the following DAE in the variables $\bar{\zeta}$ and $\bar{\alpha}$,

$$\dot{\bar{\zeta}}_t = \operatorname{Lin}\mathcal{F}\,|_{(\zeta_t^0, \alpha_t^0)}\,(\bar{\zeta}_t, \bar{\alpha}_t), \tag{31}$$
$$0 = \operatorname{Lin}\mathcal{G}\,|_{(\zeta_t^0, \alpha_t^0)}\,(\bar{\zeta}_t, \bar{\alpha}_t), \tag{32}$$
$$0 = \operatorname{Lin}\mathcal{I}\,|_{(\zeta_0^0, \zeta_T^0)}\,(\bar{\zeta}_0, \bar{\zeta}_T), \tag{33}$$

where

$$\operatorname{Lin}\mathcal{F}\,|_{(\zeta_t^0, \alpha_t^0)}\,(\bar{\zeta}_t, \bar{\alpha}_t) := \mathcal{F}'(\zeta_t^0, \alpha_t^0)(\bar{\zeta}_t, \bar{\alpha}_t), \tag{34}$$

and the analogous definitions hold for $\operatorname{Lin}\mathcal{G}$ *and* $\operatorname{Lin}\mathcal{H}$.

The technical result below will simplify the computation of the linearization of (OS). Its proof is immediate.

Lemma 4.3 (Commutation of linearization and differentiation)**.** *Given $\mathcal{G}$ and $\mathcal{F}$ as in the previous definition, it holds*

$$\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Lin}\mathcal{G} = \operatorname{Lin}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}, \quad \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{Lin}\mathcal{F} = \operatorname{Lin}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}. \tag{35}$$

4.2 Linearized optimality system

In the sequel, whenever the argument of functions A, B, B_1, etc. is omitted, assume that they are evaluated at the reference extremal $(\hat{w}, \hat{\lambda})$. Define the $m \times n-$matrix C, the $n \times n-$matrix Q and the $m \times n-$matrix M by

$$C := H_{ux}, \quad Q := H_{xx}, \quad M := B^\top Q - \dot{C} - CA. \tag{36}$$

Notice that the ith. row of matrix C is the function pf_i', for $i = 1, \dots, m$. Denote with $(z, v, \bar{\lambda} := (\bar{\beta}, q))$ the linearized variable $(x, u, \lambda = (\beta, p))$. In view of equations (21) and (36) we can write

$$\text{Lin } \Phi_t = q_t B_t + z_t^\top C_t^\top. \tag{37}$$

The linearization of system (OS) at point $(\hat{x}, \hat{u}, \hat{\lambda})$ consists of the *linearized state equation*

$$\dot{z}_t = A_t z_t + B_t v_t, \quad \text{a.e. on } [0, T], \tag{38}$$

with endpoint conditions

$$0 = D\eta(\hat{x}_0, \hat{x}_T)(z_0, z_T), \tag{39}$$

the linearized costate equation

$$-\dot{q}_t = q_t A_t + z_t^\top Q_t + v_t^\top C_t, \quad \text{a.e. on } [0, T], \tag{40}$$

with endpoint conditions

$$q_0 = -\left[z_0^\top D^2_{x_0^2} \ell + z_T^\top D^2_{x_0 x_T} \ell + \sum_{j=1}^{d_\eta} \bar{\beta}_j D_{x_0} \eta_j \right]_{(\hat{x}_0, \hat{x}_T)}, \tag{41}$$

$$q_T = \left[z_T^\top D^2_{x_T^2} \ell + z_0^\top D^2_{x_0 x_T} \ell + \sum_{j=1}^{d_\eta} \bar{\beta}_j D_{x_T} \eta_j \right]_{(\hat{x}_0, \hat{x}_T)}, \tag{42}$$

and the algebraic equations

$$0 = \text{Lin } \ddot{\Phi} = -\frac{\mathrm{d}^2}{\mathrm{d}t^2}(qB + Cz), \quad \text{a.e. on } [0, T], \tag{43}$$

$$0 = \text{Lin } \Phi_T = q_T B_T + C_T z_T, \tag{44}$$

$$0 = \text{Lin } \dot{\Phi}_0 = -\frac{\mathrm{d}}{\mathrm{d}t}(qB + Cz)_{t=0}. \tag{45}$$

Here we used equation (37) and commutation property of Lemma 4.3 to write (43) and (47). Observe that (43)-(47) and Lemma 4.3 yield

$$0 = \text{Lin } \Phi_t = q_t B_t + z_t^\top C_t^\top, \quad \text{on } [0, T], \tag{46}$$

and

$$0 = \text{Lin } \dot{\Phi}_t = -qB_1 - z^\top M^\top + v^\top(-CB + B^\top C^\top), \quad \text{a.e. on } [0, T].$$

By means of Theorem 5.2 to be stated in Section 5 afterwards we can see that the coefficient of v in previous expression vanishes, and hence,

$$0 = \text{Lin } \dot{\Phi}_t = -qB_1 - z^\top M^\top, \quad \text{on } [0, T]. \tag{47}$$

Note that both equations (46) and (47) hold everywhere on $[0,T]$ since all the involved functions are continuous in time.
Notation: denote by (LS) the set of equations (38)-(47).

Once we have computed the linearized system (LS), we can write the derivative of $\mathcal{S}$ in the direction $\bar{\nu} := \left(z_0, q_0, \bar{\beta}\right)$ as follows.

$$\mathcal{S}'(\hat{\nu})\bar{\nu} = \begin{pmatrix} D\eta(\hat{x}_0,\hat{x}_T)(z_0,z_T) \\ q_0 + \left[z_0^\top D^2_{x_0^2}\ell + z_T^\top D^2_{x_0x_T}\ell + \sum_{j=1}^{d_\eta}\bar{\beta}_j D_{x_0}\eta_j\right]_{(\hat{x}_0,\hat{x}_T)} \\ q_T - \left[z_T^\top D^2_{x_T^2}\ell + z_0^\top D^2_{x_0x_T}\ell + \sum_{j=1}^{d_\eta}\bar{\beta}_j D_{x_T}\eta_j\right]_{(\hat{x}_0,\hat{x}_T)} \\ q_T B_T + z_T^\top C_T^\top \\ q_0 B_{1,0} + z_0^\top M_0^\top \end{pmatrix}, \tag{48}$$

where (v,z,q) is the solution of (38),(40),(43) associated with the initial condition (z_0,q_0) and the multiplier $\bar{\beta}$. Thus, we get the property below.

Proposition 4.4. *$\mathcal{S}'(\hat{\nu})$ is one-to-one if the only solution of* (38)-(40),(43) *with the initial conditions $z_0 = 0,\ q_0 = 0$ and with $\bar{\beta} = 0$ is $(v,z,q) = 0$.*

5 Second Order Optimality Conditions

In this section we summarize a set of second order necessary and sufficient conditions. At the end of the section we state a sufficient condition for the local quadratic convergence of the shooting algorithm presented in Section 4. The latter is the main result of this article.

Recall the matrices C and Q defined in (36), and the space $\mathcal{W}$ given at the beginning of Section 2. Consider the quadratic mapping on $\mathcal{W}$,

$$\Omega(z,v) := \tfrac{1}{2}D^2\ell\,(z_0,z_T)^2 + \tfrac{1}{2}\int_0^T \left[z^\top Q z + 2v^\top C z\right]\mathrm{d}t. \tag{49}$$

It is a well-known result that for each $(z,v)\in\mathcal{W}$,

$$D^2\mathcal{L}\,(z,v)^2 = \Omega(z,v). \tag{50}$$

We next recall the classical second order necessary condition for optimality that states that the second variation of the Lagrangian function is nonnegative on the critical cone. In our case, the *critical cone* is given by

$$\mathcal{C} := \{(z,v)\in\mathcal{W} :\ (38)\text{-}(39)\ \text{hold}\}, \tag{51}$$

and the second order optimality condition is as follows.

Theorem 5.1 (Second order necessary optimality condition)**.** *If $\hat{w}$ is a weak minimum of (P) then*

$$\Omega(z,v)\geq 0, \quad \text{for all } (z,v)\in\mathcal{C}. \tag{52}$$

A proof of previous theorem can be found in, e.g., Levitin, Milyutin and Osmolovskii [28]. The following necessary condition is due to Goh [26] and it

is a nontrivial consequence (not immediate) of Theorem 5.1. Define first the $m \times m-$matrix

$$R := B^\top Q B - CB_1 - (CB_1)^\top - \frac{\mathrm{d}}{\mathrm{d}t}(CB). \tag{53}$$

Theorem 5.2 (Goh's Necessary Condition)**.** *If $\hat{w}$ is a weak minimum of (P), then*

$$CB \text{ is symmetric,} \tag{54}$$

and

$$R \succeq 0. \tag{55}$$

Remark 5.3. *Observe that* (54) *is equivalent to $pf_i' f_j = pf_j' f_i$, for every pair $i, j = 1, \dots, m$. These identities can be written in terms of Lie brackets as*

$$p[f_i, f_j] = 0, \quad \text{for } i, j = 1, \dots, m. \tag{56}$$

Notice that (54) *implies, in view of* (53)*, that R is symmetric. The components of matrix R can be written as*

$$R_{ij} = p[f_i, [f_j, f_0]], \tag{57}$$

and hence, its symmetry implies

$$p[f_i, [f_j, f_0]] = p[f_j, [f_i, f_0]], \quad \text{for } i, j = 1, \dots, m. \tag{58}$$

The latter expressions involving Lie brackets can be often found in the literature.

The result that we present next is due to Dmitruk [20] and is stated in terms of the coercivity of Ω in a transformed space of variables. Let us give the details of the involved transformation and the transformed second variation. Given $(z, v) \in \mathcal{W}$, define

$$y_t := \int_0^t v_s \mathrm{d}s, \tag{59}$$

$$\xi_t := z_t - B(\hat{x}_t) y_t. \tag{60}$$

This change of variables, first introduced by Goh in [29], can be perform in any linear system of differential equations, and it is known as *Goh's transformation.*

We aim to perform Goh's transformation in (49). To this end consider the spaces $\mathcal{U}_2 := L_2(0, T; \mathbb{R}^m)$ and $\mathcal{X}_2 := W_2^1(0, T; \mathbb{R}^n)$, the function $g \colon \mathbb{R}^{2n+m} \to \mathbb{R}$ defined by

$$g(\zeta_0, \zeta_T, h) := D^2 \ell \left(\zeta_0, \zeta_T + B_T h\right)^2 + h^\top C_T (2\zeta_T + B_T h), \tag{61}$$

and the quadratic mapping

$$\begin{aligned} \bar{\Omega} \colon \mathcal{X}_2 \times \mathcal{U}_2 \times \mathbb{R}^m &\to \mathbb{R} \\ (\xi, y, h) &\mapsto \tfrac{1}{2} g(\xi_0, \xi_T, h) + \tfrac{1}{2} \int_0^T \{\xi^\top Q \xi + 2 y^\top M \xi + y^\top R y\} \mathrm{d}t, \end{aligned} \tag{62}$$

where the involved matrices where introduced in (19), (36) and (53).

Proposition 5.4. *If $\hat{w}$ is a weak minimum of (P), then*

$$\Omega(z,v) = \bar{\Omega}(\xi, y, y_T), \tag{63}$$

whenever $(z,v) \in \mathcal{W}$ and $(\xi, y, y_T) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ satisfy (59)-(60)*.*

The latter result follows by integrating by parts the terms containing v in (49), and by replacing z by its expression in (60). See, e.g., Aronna et al. [30] for the detailed calculations that lead to (63).

Define the *order function* $\gamma\colon \mathbb{R}^n \times \mathcal{U}_2 \times \mathbb{R}^m \to \mathbb{R}$ as

$$\gamma(\xi_0, y, h) := |\xi_0|^2 + \int_0^T y_t^2 \mathrm{d}t + |h|^2. \tag{64}$$

We call $(\delta x, v) \in \mathcal{W}$ a *feasible variation* for $\hat{w}$ if $(\hat{x} + \delta x, \hat{u} + v)$ satisfies (2)-(3).

Definition 5.5. *We say that $\hat{w}$ satisfies the $\gamma-$growth condition in the weak sense if there exists $\rho > 0$ such that, for every sequence of feasible variations $\{(\delta x^k, v^k)\}$ converging to 0 in $\mathcal{W}$,*

$$J(\hat{u} + v^k) - J(\hat{u}) \geq \rho\gamma(\xi_0^k, y^k, y_T^k), \tag{65}$$

holds for big enough k, where $y_t^k := \int_0^t v_s^k \mathrm{d}s$, and ξ^k is given by (60)*.*

In previous definition, given that $(\delta x^k, v^k)$ is a feasible variation for each k, the sequence $\{(\delta x^k, v^k)\}$ goes to 0 in $\mathcal{W}$ if and only if $\{v^k\}$ goes to 0 in $\mathcal{U}$.

Observe that if $(z,v) \in \mathcal{W}$ satisfies (38)-(39), then $(\xi, y, h := y_T)$ given by transformation (59)-(60) verifies

$$\dot{\xi} = A\xi + B_1 y, \tag{66}$$

$$D\eta(\hat{x}_0, \hat{x}_T)(\xi_0, \xi_T + B_T h) = 0. \tag{67}$$

Set the *transformed critical cone*

$$\mathcal{P}_2 := \{(\xi, y, h) \in \mathcal{X}_2 \times \mathcal{U}_2 \times \mathbb{R}^m : (66)\text{-}(67) \text{ hold}\}. \tag{68}$$

The following is an immediate consequence of the sufficient condition established in Dmitruk [20] (or [21, Theorem 3.1]).

Theorem 5.6. *The trajectory $\hat{w}$ is a weak minimum of (P) satisfying $\gamma-$ growth condition in the weak sense if and only if* (54) *holds and there exists $\rho > 0$ such that*

$$\bar{\Omega}(\xi, y, h) \geq \rho\gamma(\xi_0, y, h), \quad \text{on } \mathcal{P}_2. \tag{69}$$

The result presented in [20] applies to a more general case having finitely many equalities and inequalities constraints on the initial and final state, and a set of multipliers consisting possibly of more than one element.

Remark 5.7. *If* (69) *holds then necessarily*

$$R \succeq \rho\, I, \tag{70}$$

where I represents the identity matrix.

Theorem 5.8. *If $\hat{w}$ is a weak minimum of (P) satisfying* (69)*, then the shooting algorithm is locally quadratically convergent.*

We present the proof of previous theorem at the end of Section 7.

Remark 5.9. *It is interesting to observe that condition* (69) *is a quite weak assumption in the sense that it is necessary for $\gamma-$growth and its corresponding relaxed condition* (52) *holds necessarily for every weak minimum.*

Remark 5.10 (Verification of (69))**.** *The sufficient condition in* (69) *can be sometimes checked analytically. On the other hand, when the initial point ξ_0 is fixed, it can be characterized by a Riccati-type equation and/or the nonexistence of a focal point as it was established in Zeidan [31]. Furthermore, under certain hypotheses, the condition* (69) *can be verified numerically as proposed in [32] by Bonnard, Caillau and Trélat (see also the survey in [13]).*

6 Corresponding LQ Problem

In this section we study the linear-quadratic problem (LQ) given by

$$\bar{\Omega}(\xi, y, h_T) \to \min, \tag{71}$$

$$(66)\text{-}(67), \tag{72}$$

$$\dot{h} = 0, \quad h_0 \text{ free}. \tag{73}$$

Here y is the control, ξ and h are the state variables. Note that if condition (69) holds then (LQ) has a unique optimal solution $(\xi, y, h) = 0$. Furthermore, recall that (69) yields (70) as it was said in Remark 5.7. In other words, (69) implies that the strengthened Legendre-Clebsch condition holds at $(\xi, y, h) = 0$. Hence, the unique local optimal solution of (LQ) is characterized by the first optimality system, that we denote afterwards by (LQS). In Section 7 we present a one-to-one linear mapping that transforms each solution of (LS) (introduced in section 4.2) into a solution of this new optimality system (LQS). Theorem 5.8 will follow.

Denote by χ and χ_h the costate variables corresponding to ξ and h, respectively; and by β^{LQ} the multiplier associated to the initial-final linearized state constraint (67). Note that the qualification hypothesis in Assumption 2.1 implies that $\{D\eta_j(\hat{x}_0, \hat{x}_T)\}_{j=1}^{d_\eta}$ are linearly independent. Hence any weak solution (ξ, y, h) of (LQ) has a unique associated multiplier $\lambda^{LQ} := (\chi, \chi_h, \beta^{LQ})$ solution of the system that we describe next. The pre-Hamiltonian of (LQ) is

$$\mathcal{H}[\lambda^{LQ}](\xi, y) := \chi(A\xi + B_1 y) + \tfrac{1}{2}(\xi^\top Q\xi + 2y^\top M\xi + y^\top Ry). \tag{74}$$

Observe that $\mathcal{H}$ does not depend on h since the latter has zero dynamics and does not appear in the running cost. The endpoint Lagrangian is given by

$$\ell^{LQ}[\lambda^{LQ}](\xi_0, \xi_T, h_T) := \tfrac{1}{2}g(\xi_0, \xi_T, h_T) + \sum_{j=1}^{d_\eta} \beta_j^{LQ} D\eta_j(\xi_0, \xi_T + B_T h_T). \tag{75}$$

The costate equation for χ is

$$-\dot{\chi} = D_\xi \mathcal{H}[\lambda^{LQ}] = \chi A + \xi^\top Q + y^\top M, \tag{76}$$

with endpoint conditions

$$\begin{aligned}\chi_0 = & -D_{\xi_0}\ell^{LQ}[\lambda^{LQ}] \\ = & -\left[\xi_0^\top D^2_{x_0^2}\ell + (\xi_T + B_T h)^\top D^2_{x_0 x_T}\ell + \textstyle\sum_{j=1}^{d_\eta}\beta_j^{LQ} D_{x_0}\eta_j\right],\end{aligned} \tag{77}$$

$$\begin{aligned}\chi_T = & D_{\xi_T}\ell^{LQ}[\lambda^{LQ}] \\ = & \xi_0^\top D^2_{x_0 x_T}\ell + (\xi_T + B_T h)^\top D^2_{x_T^2}\ell + h^\top C_T + \textstyle\sum_{j=1}^{d_\eta}\beta_j^{LQ} D_{x_T}\eta_j.\end{aligned} \tag{78}$$

For costate variable χ_h we get the equation

$$\dot\chi_h = 0, \tag{79}$$

$$\chi_{h,0} = 0, \tag{80}$$

$$\chi_{h,T} = D_h\ell^{LQ}[\lambda^{LQ}]. \tag{81}$$

Hence, $\chi_h \equiv 0$ and thus (81) yields

$$0 = \xi_0^\top D^2_{x_0 x_T}\ell B_T + (\xi_T + B_T h)^\top (D^2_{x_T^2}\ell B_T + C_T^\top) + \sum_{j=1}^{d_\eta}\beta_j^{LQ} D_{x_T}\eta_j B_T. \tag{82}$$

The stationarity with respect to the new control y implies

$$0 = D_y\mathcal{H} = \chi B_1 + \xi^\top M^\top + y^\top R. \tag{83}$$

Notation: Denote by (LQS) the set of equations consisting of (66)-(67), (73),(76)-(78),(82) and (83), i.e. (LQS) is the system

$$\left\{\begin{array}{l} \dot\xi = A\xi + B_1 y, \\ D\eta(\hat x_0, \hat x_T)(\xi_0, \xi_T + B_T h) = 0, \\ \dot h = 0, \\ -\dot\chi = D_\xi\mathcal{H}[\lambda^{LQ}] = \chi A + \xi^\top Q + y^\top M, \\ \chi_0 = -\left[\xi_0^\top D^2_{x_0^2}\ell + (\xi_T + B_T h)^\top D^2_{x_0 x_T}\ell + \displaystyle\sum_{j=1}^{d_\eta}\beta_j^{LQ} D_{x_0}\eta_j\right], \\ \chi_T = \xi_0^\top D^2_{x_0 x_T}\ell + (\xi_T + B_T h)^\top D^2_{x_T^2}\ell + h^\top C_T + \displaystyle\sum_{j=1}^{d_\eta}\beta_j^{LQ} D_{x_T}\eta_j, \\ 0 = \xi_0^\top D^2_{x_0 x_T}\ell B_T + (\xi_T + B_T h)^\top (D^2_{x_T^2}\ell B_T + C_T^\top) + \displaystyle\sum_{j=1}^{d_\eta}\beta_j^{LQ} D_{x_T}\eta_j B_T, \\ 0 = \chi B_1 + \xi^\top M^\top + y^\top R. \end{array}\right. \tag{LQS}$$

Notice that (LQS) is a first order optimality system for problem (71)-(73).

7 The Transformation

In this section we show how to transform a solution of (LS) into a solution of (LQS) via a one-to-one linear mapping. Given $(z, v, q, \bar\beta) \in \mathcal{X}\times\mathcal{U}\times\mathcal{X}_*\times\mathbb{R}^{d_\eta,*}$,

define

$$y_t := \int_0^t v_s ds,\ \xi := z - By,\ \chi := q + y^\top C,\ \chi_h := 0,\ h := y_T,\ \beta_j^{LQ} := \bar{\beta}_j. \quad (84)$$

The next lemma shows that the point $(\xi, y, h, \chi, \chi_h, \beta^{LQ})$ is solution of (LQS) provided that $(z, v, q, \bar{\beta})$ is solution of (LS).

Lemma 7.1. *The one-to-one linear mapping defined by* (84) *converts each solution of (LS) into a solution of (LQS).*

Proof. Let $(z, v, q, \bar{\beta})$ be a solution of (LS), and set $(\xi, y, \chi, \beta^{LQ})$ by (84).
Part I. We shall prove that $(\xi, y, \chi, \beta^{LQ})$ satisfies conditions (66) and (67). Equation (66) follows by differentiating expression of ξ in (84), and equation (67) follows from (39).
Part II. We shall prove that $(\xi, y, \chi, \beta^{LQ})$ verifies (76)-(78) and (82). Differentiate χ in (84), use equations (40) and (84), recall definition of M in (36) and obtain

$$\begin{array}{rcl} -\dot{\chi} = & -\dot{q} - v^\top C - y^\top \dot{C} \\ = & qA + z^\top Q - y^\top \dot{C} \\ = & \chi A + \xi^\top Q + y^\top(-CA + B^\top Q - \dot{C}) \\ = & \chi A + \xi^\top Q + y^\top M. \end{array} \quad (85)$$

Hence (76) holds. Equations (77) and (78) follow from (41) and (42). Combine (42) and (44) to get

$$\begin{array}{rcl} 0 = & q_T B_T + z_T^\top C_T^\top \\ = & \left[z_T^\top D^2_{x_T^2}\ell + z_0^\top D^2_{x_0 x_T}\ell + \sum_{j=1}^{d_\eta} \bar{\beta}_j D_{x_T}\eta_j\right]_{(\hat{x}_0, \hat{x}_T)} B_T + z_T^\top C_T^\top. \end{array} \quad (86)$$

Performing transformation (84) in the previous equation yields (82).
Part III. We shall prove that (83) holds. Differentiating (46) we get

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Lin}\ \Phi = \frac{\mathrm{d}}{\mathrm{d}t}(qB + z^\top C^\top). \quad (87)$$

Consequently, by (38) and (40),

$$0 = -(qA + z^\top Q + v^\top C)B + q\dot{B} + (z^\top A^\top + v^\top B^\top)C^\top + z^\top \dot{C}^\top, \quad (88)$$

where the coefficient of v vanishes in view of (54). Recall (20) and (36). Performing transformation (84) and regrouping the terms we get from (88),

$$0 = -\chi B_1 - \xi^\top M^\top + y^\top(CB_1 - B^\top Q B + B^\top A^\top C^\top + B^\top \dot{C}^\top). \quad (89)$$

Equation (83) follows from (53) and condition (54).

Parts I, II and III show that $(\xi, y, \chi, \beta^{LQ})$ is a solution of (LQS), and hence the result follows.

□

Remark 7.2. *Observe that the unique assumption we needed in previous proof was Goh's condition* (54) *that follows from the weak optimality of* $\hat{w}$.

Proof. [of Theorem 5.8] We shall prove that (69) implies that $\mathcal{S}'(\hat{\nu})$ is one-to-one. Take $(z, v, q, \bar{\beta})$ a solution of (LS), and let $(\xi, y, \chi, \chi_h, \beta^{LQ})$ be defined by (84), that we know by Lemma 7.1 is solution of (LQS). As it has been already pointed out at the beginning of Section 6, condition (69) implies that the unique solution of (LQS) is 0. Hence $(\xi, y, \chi, \chi_h, \beta^{LQ}) = 0$ and thus $(z, v, q, \bar{\beta}) = 0$. Conclude that the unique solution of (LS) is 0. The latter assertion implies, in view of Proposition 4.4, that $\mathcal{S}'(\hat{\nu})$ is one-to-one. The result follows from Proposition 4.1. □

8 Control Constrained Case

In this section we add the following bounds to the control variables

$$0 \leq u_{i,t} \leq 1, \quad \text{for a.a. } t \in [0, T], \text{ for } i = 1, \ldots, m. \tag{90}$$

Denote with (CP) the problem given by (1)-(3) and (90).

Definition 8.1. *A feasible trajectory $\hat{w} \in \mathcal{W}$ is a Pontryagin minimum of (CP) if for any positive N there exists $\varepsilon_N > 0$ such that $\hat{w}$ is a minimum in the set of feasible trajectories $w = (x, u) \in \mathcal{W}$ satisfying*

$$\|x - \hat{x}\|_\infty < \varepsilon_N, \ \|u - \hat{u}\|_1 < \varepsilon_N, \ \|u - \hat{u}\|_\infty < N.$$

Given $i = 1, \ldots, m$, we say that $\hat{u}_i$ has a *bang arc* in $(a, b) \subset (0, T)$ if $\hat{u}_{i,t} = 0$ a.e. on (a, b) or $\hat{u}_{i,t} = 1$ a.e. on (a, b), and it has a *singular arc* if $0 < \hat{u}_{i,t} < 1$ a.e. on (a, b).

Assumption 8.2. *Each component $\hat{u}_i$ is a finite concatenation of bang and singular arcs.*

A time $t \in (0, T)$ is called *switching time* if there exists an index $1 \leq i \leq m$ such that $\hat{u}_i$ switches at time t from singular to bang, or vice versa, or from one bound in (90) to the other.

Remark 8.3. *Assumption 8.2 rules out the solutions having an infinite number of swit- chings in a bounded interval. This behavior is usually known as Fuller's phenomenon (see Fuller [33]). Many examples can be encountered satisfying Assumption 8.2 as is the case of the three problems presented in Section 10.*

With the purpose of solving (CP) numerically we assume that the structure of the concatenation of bang and singular arcs of the optimal solution $\hat{w}$ and an approximation of its switching times are known. This initial guess can be obtained, for instance, by solving the nonlinear problem resulting from the discretization of the optimality conditions or by a continuation method. See Betts [34] or Biegler [35] for a detailed survey and description of numerical methods for nonlinear programming problems. For the continuation method the reader is referred to Martinon [9].

This section is organized as follows. From (CP) and the known structure of $\hat{u}$ and its switching times we create a new problem that we denote by (TP). Afterwards we prove that we can transform $\hat{w}$ into a weak solution $\hat{W}$ of (TP). Finally we conclude that if $\hat{W}$ satisfies the coercivity condition (69), then the shooting method for problem (TP) converges locally quadratically. In practice,

the procedure will be as follows: obtain somehow the structure of the optimal solution of (CP), create problem (TP), solve (TP) numerically obtaining $\hat{W}$, and finally transform $\hat{W}$ to find $\hat{w}$.

Next we present the transformed problem.

Assumption 8.4. *Assume that each time a control $\hat{u}_i$ switches from bang to singular or vice versa, there is a discontinuity of first kind.*

Here, by *discontinuity of first kind* we mean that each component of $\hat{u}$ has a finite nonzero jump at the switching times, and the left and right limits exist.

By Assumption 8.2 the set of switching times is finite. Consider the partition of $[0, T]$ induced by the switching times:

$$\{0 =: \hat{T}_0 < \hat{T}_1 < \ldots < \hat{T}_{N-1} < \hat{T}_N := T\}. \tag{91}$$

Set $\hat{I}_k := [\hat{T}_{k-1}, \hat{T}_k]$, and define for $k = 1, \ldots, N$,

$$S_k := \{1 \leq i \leq m : \ \hat{u}_i \text{ is singular on } \hat{I}_k\}, \tag{92}$$

$$E_k := \{1 \leq i \leq m : \ \hat{u}_i = 0 \text{ a.e. on } \hat{I}_k\}, \tag{93}$$

$$N_k := \{1 \leq i \leq m : \ \hat{u}_i = 1 \text{ a.e. on } \hat{I}_k\}. \tag{94}$$

Clearly $S_k \cup E_k \cup N_k = \{1, \ldots, m\}$.

Assumption 8.5. *For each $k = 1, \ldots, N$, denote by u_{S_k} the vector with components u_i with $i \in S_k$. Assume that the strengthened generalized Legendre-Clebsch condition holds on $\hat{I}_k$, i.e.*

$$-\frac{\partial}{\partial u_{S_k}} \ddot{H}_{u_{S_k}} \succ 0, \quad \text{on } \hat{I}_k. \tag{95}$$

Hence, u_{S_k} can be retrieved from equation

$$\ddot{H}_{u_{S_k}} = 0, \tag{96}$$

since the latter is affine on u_{S_k} as it has been already pointed out in Section 3. Observe that the expression obtained from (96) involves only the state variable $\hat{x}$ and the corresponding adjoint state $\hat{p}$. Hence, it results that $\hat{u}_{S_k}$ is continuous on $\hat{I}_k$ with finite limits at the endpoints of this interval. As the components $\hat{u}_i$ with $i \notin S_k$ are either identically 1 or 0, we conclude that

$$\hat{u} \text{ is continuous on } \hat{I}_k. \tag{97}$$

By Assumption 8.4 and condition (97) (derived from Assumption 8.5) we get that there exists $\rho > 0$ such that

$$\rho < \hat{u}_{i,t} < 1 - \rho, \quad \text{a.e. on } \hat{I}_k, \text{ for } k = 1, \ldots, N, \ i \in S_k. \tag{98}$$

Next we present a new control problem obtained in the following way. For each $k = 1, \ldots, N$, we perform the change of time variable that converts the interval $\hat{I}_k$ into $[0, 1]$, afterwards we fix the bang control variables to their bounds and finally, we associate a free control variable to each index in S_k. More precisely, consider for $k = 1, \ldots, N$ the control variables $u_i^k \in L_\infty(0, 1; \mathbb{R})$, with $i \in S_k$, and the state variables $x^k \in W_\infty^1(0, 1; \mathbb{R}^n)$. Let the constants $T_k \in \mathbb{R}$, for $k =$

$1, \ldots, N-1$, which will be considered as state variables of zero-dynamics. Set $T_0 := 0$, $T_N := T$ and define the problem on the interval $[0,1]$

$$\varphi_0(x_0^1, x_1^N) \to \min, \tag{99}$$

$$\dot{x}^k = (T_k - T_{k-1}) \left(\sum_{i \in N_k \cup \{0\}} f_i(x^k) + \sum_{i \in S_k} u_i^k f_i(x^k) \right), \quad k = 1, \ldots, N, \tag{100}$$

$$\dot{T}_k = 0, \quad k = 1, \ldots, N-1, \tag{101}$$

$$\eta(x_0^1, x_1^N) = 0, \tag{102}$$

$$x_1^k = x_0^{k+1}, \quad k = 1, \ldots, N-1. \tag{103}$$

Denote by (TP) the problem consisting of equations (99)-(103). The link between the original problem (CP) and the transformed one (TP) is given in Lemma 8.6 below. Set for each $k = 1, \ldots, N$:

$$\hat{x}_s^k := \hat{x}(\hat{T}_{k-1} + (\hat{T}_k - \hat{T}_{k-1})s), \quad \text{for } s \in [0,1], \tag{104}$$

$$\hat{u}_{i,s}^k := \hat{u}_i(\hat{T}_{k-1} + (\hat{T}_k - \hat{T}_{k-1})s), \quad \text{for } i \in S_k, \text{ a.a. } s \in [0,1]. \tag{105}$$

Set

$$\hat{W} := ((\hat{x}^k)_{k=1}^N, (\hat{u}_i^k)_{k=1, i \in S_k}^N, (\hat{T}_k)_{k=1}^{N-1}). \tag{106}$$

Lemma 8.6. *If $\hat{w}$ is a Pontryagin minimum of (CP), then $\hat{W}$ is a weak solution of (TP).*

Proof. The idea of the proof is to derive the weak optimality of $\hat{W}$ from the Pontryagin optimality of $\hat{w}$ and condition (98). Since $\hat{w}$ is a Pontryagin minimum for (CP), there exists $\varepsilon > 0$ such that $\hat{w}$ is a minimum in the set of feasible trajectories $w = (x, u)$ satisfying

$$\|x - \hat{x}\|_\infty < \varepsilon, \quad \|u - \hat{u}\|_1 < \varepsilon, \quad \|u - \hat{u}\|_\infty < 1. \tag{107}$$

Consider $\bar{\delta}, \bar{\varepsilon} > 0$, and a feasible solution $((x^k), (u_i^k), (T_k))$ for (TP) such that

$$|T_k - \hat{T}_k| \le \bar{\delta}, \quad \|u_i^k - \hat{u}_i^k\|_\infty < \bar{\varepsilon}, \quad \text{for all } k = 1, \ldots, N. \tag{108}$$

We shall relate ε in (107) with $\bar{\delta}$ and $\bar{\varepsilon}$ in (108). Let $k = 1, \ldots, N$. Denote $I_k := (T_{k-1}, T_k)$, and define for each $i = 1, \ldots, m$:

$$u_{i,t} := \begin{cases} 0, & \text{if } t \in I_k \text{ and } i \in E_k, \\ u_i^k \left(\frac{t - T_{k-1}}{T_k - T_{k-1}} \right), & \text{if } t \in I_k \text{ and } i \in S_k, \\ 1, & \text{if } t \in I_k \text{ and } i \in N_k. \end{cases} \tag{109}$$

Let x be the solution of (2) associated to u and having $x_0 = x_0^1$. We shall prove that (x, u) is feasible for the original problem (CP). Observe that condition (103) implies that $x_t = x^k \left(\frac{t - T_{k-1}}{T_k - T_{k-1}} \right)$ when $t \in I_k$, and thus $x_1 = x_1^N$. It follows that (3) holds. We shall check condition (90). For $i \in E_k \cup N_k$, it follows from the definition in (109). Consider now $i \in S_k$. Since (98) holds, by (105) we get

$$\rho < \hat{u}_{i,s}^k < 1 - \rho, \quad \text{a.e. on } (0,1). \tag{110}$$

Thus, by (108) and if $\bar{\varepsilon} < \rho$, we get $0 < u^k_{i,s} < 1$ a.e. on $(0,1)$. This yields

$$0 < u_{i,t} < 1, \quad \text{a.e. on } I_k, \tag{111}$$

and thus the feasibility of (x,u) for (CP).

We now estimate $\|u-\hat{u}\|_1$. For $k=1,\dots,N$ and $i \in S_k$,

$$\begin{array}{rl} \int_{I_k\cap\hat{I}_k} |u_{i,t}-\hat{u}_{i,t}|\mathrm{d}t \leq & \int_{I_k\cap\hat{I}_k} \left| u^k_i\left(\frac{t-T_{k-1}}{T_k-T_{k-1}}\right) - \hat{u}^k_i\left(\frac{t-T_{k-1}}{T_k-T_{k-1}}\right)\right| \mathrm{d}t \\ & + \int_{I_k\cap\hat{I}_k} \left| \hat{u}^k_i\left(\frac{t-T_{k-1}}{T_k-T_{k-1}}\right) - \hat{u}^k_i\left(\frac{t-\hat{T}_{k-1}}{\hat{T}_k-\hat{T}_{k-1}}\right)\right| \mathrm{d}t. \end{array} \tag{112}$$

Note that by Assumption 8.4 and condition (97), each $\hat{u}^k_i$ is uniformly continuous on $\hat{I}_k$, and thus there exists $\theta_{ki} > 0$ such that if $|s-s'| < \theta_{ki}$ then $|\hat{u}^k_{i,s} - \hat{u}^k_{i,s'}| < \bar{\varepsilon}$. Set $\bar{\theta} := \min \theta_{ki} > 0$. Consider then $\bar{\delta}$ such that if $|T_k - \hat{T}_k| < \bar{\delta}$, then $\left|\frac{t-T_{k-1}}{T_k-T_{k-1}} - \frac{t-\hat{T}_{k-1}}{\hat{T}_k-\hat{T}_{k-1}}\right| < \bar{\theta}$. From (108) and (112) we get

$$\int_{I_k\cap\hat{I}_k} |u_{i,t}-\hat{u}_{i,t}|\mathrm{d}t < 2\bar{\varepsilon}\,\mathrm{meas}\,(I_k\cap\hat{I}_k). \tag{113}$$

Assume, w.l.o.g., that $T_k < \hat{T}_k$ and note that

$$\int_{T_k}^{\hat{T}_k} |u_{i,t}-\hat{u}_{i,t}|\mathrm{d}t \leq \int_{T_k}^{\hat{T}_k} \left| u^k_i\left(\frac{t-T_{k-1}}{T_k-T_{k-1}}\right) - \hat{u}^k_i\left(\frac{t-\hat{T}_{k-1}}{\hat{T}_k-\hat{T}_{k-1}}\right)\right| \mathrm{d}t < \bar{\delta}\,\bar{\varepsilon}, \tag{114}$$

where we used (108) in the last inequality. From (113) and (114) we get $\|u_i - \hat{u}_i\|_1 < \bar{\varepsilon}(2T + (N-1)\bar{\delta})$. Thus $\|u-\hat{u}\|_1 < \varepsilon$ if

$$\bar{\varepsilon}(2T + (N-1)\bar{\delta}) < \varepsilon/m. \tag{115}$$

We conclude from (107) that $((x^k),(u^k_i),(T_k))$ is a minimum on the set of feasible points satisfying (108) and (115). Thus $\hat{W}$ is a weak solution of (TP), as it was to be proved. □

We shall next propose a shooting function associated to (TP). The pre-Hamilto- nian of the latter is

$$\tilde{H} := \sum_{k=1}^{N} (T_k - T_{k-1}) H^k, \tag{116}$$

where, denoting by p^k the costate variable associated to x^k,

$$H^k := p^k \left(\sum_{i\in N_k\cup\{0\}} f_i(x^k) + \sum_{i\in S_k} u^k_i f_i(x^k) \right). \tag{117}$$

Observe that Assumption 8.5 made on $\hat{u}$ yields

$$-\frac{\partial}{\partial u}\ddot{H}_u \succ 0, \quad \text{on } [0,1], \tag{118}$$

i.e. the strengthened generalized Legendre-Clebsch condition holds in problem (TP) at $\hat{w}$. Hence we can define the shooting function for (TP) as it was done in Section 4 for (P).

The endpoint Lagrangian is

$$\tilde{\ell} := \varphi_0(x_0^1, x_1^N) + \sum_{j=1}^{d_\eta} \beta_j \eta_j(x_0^1, x_1^N) + \sum_{k=1}^{N-1} \theta_k (x_1^k - x_0^{k+1}). \tag{119}$$

The costate equation for p^k is given by

$$\dot{p}^k = -(T_k - T_{k-1}) D_{x^k} H^k, \tag{120}$$

with endpoint conditions

$$p_0^1 = -D_{x_0^1}\tilde{\ell} = -D_{x_0^1}\varphi_0 - \sum_{j=1}^{d_\eta} \beta_j D_{x_0^1}\eta_j, \tag{121}$$

$$\begin{array}{ll} p_1^k & = \theta^k, \quad \text{for } k = 1, \dots, N-1, \\ p_0^k & = \theta^{k-1}, \quad \text{for } k = 2, \dots, N, \end{array} \tag{122}$$

$$p_1^N = D_{x_1^N}\tilde{\ell} = D_{x_1^N}\varphi_0 + \sum_{j=1}^{d_\eta} \beta_j D_{x_1^N}\eta_j. \tag{123}$$

For the costate variables p^{T_k} associated with T_k we get the equations

$$\dot{p}^{T_k} = -H^k + H^{k+1}, \quad p_0^{T_k} = 0, \quad p_1^{T_k} = 0, \quad \text{for } k = 1, \dots, N-1, . \tag{124}$$

Remark 8.7. *We can sum up the conditions in* (124) *integrating the first one and obtaining* $\int_0^1 (H^{k+1} - H^k)\mathrm{d}t = 0$, *and hence, since* H^k *is constant on the optimal trajectory, we get the equivalent condition*

$$H_1^k = H_0^{k+1}, \quad \text{for } k = 1, \dots, N-1. \tag{125}$$

So we can remove the shooting variable p^{T_k} *and keep the continuity condition on the pre-Hamiltonian.*

Observe that (103) and (122) imply the continuity of the two functions obtained by concatenating the states and the costates, i.e. the continuity of X and P defined by

$$X_0 := x_0^1, \ X_s := x^k(s - (k-1)), \quad \text{for } s \in (k-1, k], \ k = 1, \dots, N, \tag{126}$$

$$P_0 := p_0^1, \ P_s := p^k(s - (k-1)), \quad \text{for } s \in (k-1, k], \ k = 1, \dots, N. \tag{127}$$

Thus, while iterating the shooting method, we can either include the conditions (103) and (122) in the definition of the shooting function or integrate the differential equations for x^k and p^k from the values x_1^{k-1} and p_1^{k-1} previously obtained. The latter option reduces the number of variables and hence the size of the problem, but is less stable. We shall present below the shooting function for the more stable case. For this end define the $n \times n-$matrix

$$A^k := \sum_{i \in N_k \cup \{0\}} f_i'(\hat{x}^k) + \sum_{i \in S_k} \hat{u}_i^k f_i'(\hat{x}^k), \tag{128}$$

the $n \times |S_k|-$matrix B^k with columns $f_i(\hat{x}^k)$ with $i \in S_k$, and

$$B_1^k := A^k B^k - \frac{\mathrm{d}}{\mathrm{d}t} B^k. \tag{129}$$

We shall denote by $g_i(x^k,u^k)$ the ith. column of B_1^k for each i in S_k. Here u^k is the $|S_k|-$dimensional vector of components u_i^k. The resulting shooting function for (TP) is given by

$$\begin{aligned}\mathcal{S}\colon \mathbb{R}^{Nn+N-1}\times\mathbb{R}^{Nn+d_\eta,*} &\to \mathbb{R}^{d_\eta+(N-1)n}\times\mathbb{R}^{(N+1)n+N-1+2\sum|S_k|,*},\\ \big((x_0^k),(T_k),(p_0^k),\beta\big)=:\nu &\mapsto \mathcal{S}(\nu):=\begin{pmatrix}\eta(x_0^1,x_1^N)\\ (x_1^k-x_0^{k+1})_{k=1,\dots,N-1}\\ p_0^1+D_{x_0^1}\tilde\ell[\lambda](x_0^1,x_1^N)\\ (p_1^k-p_0^{k+1})_{k=1,\dots,N-1}\\ p_1^N-D_{x_1^N}\tilde\ell[\lambda](x_0^1,x_1^N)\\ (H_1^k-H_0^{k+1})_{k=1,\dots,N-1}\\ (p_0^k f_i(x_0^k))_{k=1,\dots,N,\ i\in S_k}\\ (p_0^k g_i(x_0^k,u_0^k))_{k=1,\dots,N,\ i\in S_k}\end{pmatrix}.\end{aligned} \tag{130}$$

Here we put both conditions $\tilde H_u=0$ and $\dot{\tilde H}_u=0$ at the beginning of the interval since we have already pointed out in Remark 3.2 that all the possible choices were equivalent.

Since problem (TP) has the same structure than problem (P) in section 2, i.e. they both have free control variable (initial-final constraints), we can apply Theorem 5.8 and obtain the analogous result below.

Theorem 8.8. *Assume that $\hat w$ is a Pontryagin minimum of (CP) such that $\hat W$ defined in* (106) *satisfies condition* (69) *for problem (TP). Then the shooting algorithm for (TP) is locally quadratically convergent.*

Remark 8.9. *Once system* (130) *is obtained, observe that two numerical implementations can be done: one integrating each variable on the interval* $[0,1]$ *and the other one, going back to the original interval* $[0,T]$*, and using implicitly the continuity conditions* (103)*,* (122) *and* (125) *at each switching time. The latter implementation is done in the numerical tests of Section 10 below. In this case, the sensibility with respect to the switching times is obtained from the derivative of the shooting function.*

8.1 Reduced Systems

In some cases we can show that some of the conditions imposed to the shooting function in (130) are redundant. Hence, they can be removed from the formulation yielding a smaller system that we will refer as *reduced system* and which is associated to a *reduced shooting function.*

Recall that when defining $\mathcal{S}$ we are implicitly imposing that $\ddot{\tilde H}_u\equiv 0$. The latter condition together with $\dot{\tilde H}_{u,0}=\tilde H_{u,1}=0$, both included in the definition of $\mathcal{S}$, imply that $\dot{\tilde H}_u\equiv\tilde H_u\equiv 0$. Hence,

$$p_1^k f_i(x_1^k)=p_1^k g_i(x_1^k,u_1^k)=0,\quad \text{for } k=1,\dots,N,\ i\in S_k, \tag{131}$$

and, in view of the continuity conditions (103) and (122),

$$p_0^{k+1} f_i(x_0^{k+1})=p_0^{k+1} g_i(x_0^{k+1},u_0^{k+1})=0,\quad \text{for } k=1,\dots,N-1,\ i\in S_k. \tag{132}$$

Therefore, if a component of the control is singular on I_k and remains being singular on I_{k+1}, then there is no need to impose the boundary conditions on $\tilde{H}_u$ and $\dot{\tilde{H}}_u$ since they are a consequence of the continuity conditions and the implicit equation $\ddot{\tilde{H}}_u \equiv 0$.

Observe now that from (117), (130) and previous two equations (131) and (132) we obtain,

$$H_1^k = p_1^k \sum_{N_k \cup \{0\}} f_i(x_1^k) = p_0^{k+1} \sum_{N_k \cup \{0\} \setminus S_{k+1}} f_i(x_0^{k+1}). \tag{133}$$

On the other hand,

$$H_0^{k+1} = p_0^{k+1} \sum_{N_{k+1} \cup \{0\} \setminus S_k} f_i(x_0^{k+1}). \tag{134}$$

Thus, $H_1^k = H_0^{k+1}$ if $N_k \cup \{0\} \setminus S_{k+1} = N_{k+1} \cup \{0\} \setminus S_k$. The latter equality holds if and only if at instant T_k all the switchings are either bang-to-singular or singular-to-bang.

Definition 8.10 (Reduced shooting function)**.** *We call reduced shooting function and we denote it by $\mathcal{S}^r$ the function obtained from $\mathcal{S}$ defined in* (130) *by removing the condition $H_1^k = H_0^{k+1}$ whenever all the switchings occurring at T_k are either bang-to-singular or singular-to-bang, and removing*

$$p_0^k f_i(x_0^k) = 0, \quad p_0^k g_i(x_0^k, u_0^k) = 0, \tag{135}$$

for $k = 2, \ldots, N$ and $i \in S_{k-1} \cap S_k$.

8.2 Square Systems

The reduced system above-presented can occasionally result *square*, in the sense that the reduced function $\mathcal{S}^r$ has as many variables as outputs. This situation occurs, e.g., in problems 1 and 3 of Section 10. The fact that the reduced system turns out to be square is a consequence of the structure of the optimal solution. In general, the optimal solution $\hat{u}$ yields a square reduced system if and only if each singular arc is in the interior of $[0, T]$ and at each switching time only one control component switches. This can be interpreted as follows: each singular arc contributes to the formulation with two inputs that are its entry and exit times, and with two outputs that correspond to $p_0^k f_i(x_0^k) = g_i(x_0^k, u_0^k) = 0$, being I_k the first interval where the component is singular and i the index of the analyzed component. On the other hand, whenever a bang-to-bang transition occurs, it contributes to the formulation with one input for the switching time and one output associated to the continuity of the pre-Hamiltonian (which is sometimes expressed as a zero of the switching function).

9 Stability under Data Perturbation

In this section we investigate the stability of the optimal solution under data perturbation. We shall prove that, under condition (69), the solution is stable under small perturbations of the data functions φ_0, f_i and η. Assume for this

stability analysis that the shooting system of the studied problem can be reduced to a square one. We gave a description of this situation in Subsection 8.2. Even if the above-mentioned square systems appear in control constrained problems, we start this section by establishing a stability result of the optimal solution for an unconstrained problem. Afterwards, in Subsection 9.2, we apply the latter result to problem (TP) and this way we obtain a stability result for the control constrained problem (CP).

9.1 Unconstrained control case

Consider then problem (P) presented in Section 2, and the family of problems depending on the real parameter μ given by:

$$\begin{aligned} &\varphi_0^\mu(x_0, x_T) \to \min, \\ &\dot{x}_t = \sum_{i=0}^{m} u_{i,t} f_i^\mu(x_t), \quad \text{for } t \in (0,T), \\ &\eta^\mu(x_0, x_T) = 0. \end{aligned} \tag{P_μ}$$

Assume that $\varphi_0^\mu : \mathbb{R}^{2n+1} \to \mathbb{R}$ and $\eta^\mu : \mathbb{R}^{2n+1} \to \mathbb{R}^{d_\eta}$ have Lipschitz-continuous second derivatives in the variable (x_0, x_T) and continuously differentiable with respect to μ, and $f_i^\mu : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is twice continuously differentiable with respect to x and continuously differentiable with respect to the parameter μ. In this formulation, the problem (P_0) associated to $\mu = 0$ coincides with (P), i.e. $\varphi_0^0 = \varphi_0$, $f_i^0 = f_i$ for $i = 0, \dots, m$ and $\eta^0 = \eta$. Recall (69) in Theorem 5.6, and write the analogous condition for (P_μ) as follows:

$$\bar{\Omega}^\mu(\xi, y, h) \geq \rho\gamma(\xi_0, y, h), \quad \text{on } \mathcal{P}_2^\mu, \tag{136}$$

where $\bar{\Omega}^\mu$ and $\mathcal{P}_2^\mu$ are the second variation and critical cone associated to (P_μ), respectively. Let $\mathcal{S}^\mu$ be the shooting function for (P_μ). Thus, we can write

$$\begin{aligned} \mathcal{S}^\mu : \quad \mathbb{R}^M \times \mathbb{R} \quad &\to \mathbb{R}^M, \\ (\,\nu\,,\,\mu\,) \quad &\mapsto \mathcal{S}^\mu(\nu), \end{aligned} \tag{137}$$

where we indicate with M the dimension of the domain of $\mathcal{S}$. The following stability result will be established.

Theorem 9.1 (Stability of the optimal solution)**.** *Assume that the shooting system generated by problem (P) is square and let $\hat{w}$ be a solution satisfying the uniform positivity condition* (69)*. Then there exists a neighborhood $\mathcal{J} \subset \mathbb{R}$ of 0, and a continuous differentiable mapping $\mu \mapsto w^\mu = (x^\mu, u^\mu)$, from $\mathcal{J}$ to $\mathcal{W}$, where w^μ is a weak solution for* (P_μ)*. Furthermore, w^μ verifies the uniform positivity* (136)*. Therefore, in view of Theorems 5.6 and 5.8, the $\gamma-$ growth holds, and the shooting algorithm for (P^μ) is locally quadratically convergent.*

Let us start showing the following stability result for the family of shooting functions $\{\mathcal{S}^\mu\}$.

Lemma 9.2. *Under the hypotheses of Theorem 9.1, there exists a neighborhood $\mathcal{I} \subset \mathbb{R}$ of 0 and a continuous differentiable mapping $\mu \mapsto \nu^\mu = (x_0^\mu, p_0^\mu, \beta^\mu)$, from $\mathcal{I}$ to $\mathbb{R}^M$, such that $\mathcal{S}^\mu(\nu^\mu) = 0$. Furthermore, the solutions (x^μ, u^μ, p^μ) of*

(2)-(8)-(15) *with initial condition* (x_0^μ, p_0^μ) *and associated multiplier* β^μ *provide a family of feasible trajectories* $w^\mu := (x^\mu, u^\mu)$ *verifying*

$$\|x^\mu - \hat{x}\|_\infty + \|u^\mu - \hat{u}\|_\infty + \|p^\mu - \hat{p}\|_\infty + |\beta^\mu - \hat{\beta}| = \mathcal{O}(\mu). \tag{138}$$

Proof. Since (69) holds, the result in Theorem 5.8 yields the non-singularity of the square matrix $D_\nu S^0(\hat{\nu})$. Hence, the Implicit Function Theorem is applicable and we can then guarantee the existence of a neighborhood $\mathcal{B} \subset \mathbb{R}^M$ of $\hat{\nu}$, a neighborhood $\mathcal{I} \subset \mathbb{R}$ of 0, and a continuously differentiable function $\Gamma : \mathcal{I} \to \mathcal{B}$ such that

$$\mathcal{S}^\mu(\Gamma(\mu)) = 0, \quad \text{for all } \mu \in \mathcal{I}. \tag{139}$$

Finally, write $\nu^\mu := \Gamma(\mu)$ and use the continuity of $D\Gamma$ on $\mathcal{I}$ to get the first part of the statement.

The feasibility of w^μ holds since equation (139) is verified. Finally, the estimation (138) follows from the stability of the system of differential equation provided by the shooting method. □

Once we obtained the existence of this w^μ feasible for (P^μ), we may wonder whether it is locally optimal. For this aim, we shall investigate the stability of the sufficient condition (69). Denote by $\bar{\Omega}^\mu$ and $\mathcal{P}_2^\mu$ the quadratic mapping and critical cone related to (P^μ), respectively. Given that all the functions involved in $\bar{\Omega}^\mu$ are continuously differentiable with respect to μ, the mapping $\bar{\Omega}^\mu$ itself is continuously differentiable with respect to μ. For the perturbed cone we get the following approximation result.

Lemma 9.3. *Assume the same hypotheses as in Theorem 9.1. Take* $\mu \in \mathcal{I}$ *and* $(\xi^\mu, y^\mu, h^\mu) \in \mathcal{P}_2^\mu$. *Then there exists* $(\xi, y, h) \in \mathcal{P}_2$ *such that*

$$|\xi_0^\mu - \xi_0| + \|y^\mu - y\|_2 + |h^\mu - h| = \mathcal{O}(\mu). \tag{140}$$

The definition below will be useful in the proof of previous Lemma.

Definition 9.4. *Define the function* $\bar{\eta} : \mathcal{U} \times \mathbb{R}^n \to \mathbb{R}^{d_\eta}$, *given by*

$$\bar{\eta}(u, x_0) := \eta(x_0, x_T), \tag{141}$$

where x *is the solution of* (2) *associated to* (u, x_0).

Proof. [of Lemma 9.3] Recall that $D\bar{\eta}(\hat{u}, \hat{x}_0)$ is onto by Assumption 2.1. Call back the definition of the critical cone $\mathcal{C}$ given in (51), and notice that we can rewrite it as $\mathcal{C} = \{(z, v) \in \mathcal{W} : \mathcal{G}(z, v) = 0\} = \operatorname{Ker} \mathcal{G}$, with $\mathcal{G}(z, v) := D\eta(\hat{x}_0, \hat{x}_T)(z_0, z_T)$ being an onto linear application from $\mathcal{W}$ to $\mathbb{R}^{d_\eta}$. In view of Goh's Transformation (59)-(60),

$$D\eta(\hat{x}_0, \hat{x}_T)(z_0, z_T) = D\eta(\hat{x}_0, \hat{x}_T)(\xi_0, \xi_T + B_T y_T), \tag{142}$$

for $(z, v) \in \mathcal{W}$ and (ξ, y) being its corresponding transformed direction. Thus, the cone $\mathcal{P}_2$ can be written as $\mathcal{P}_2 = \{\zeta \in \mathcal{H} : \mathcal{K}(\zeta) = 0\} = \operatorname{Ker} \mathcal{K}$, with $\zeta := (\xi, y, h)$, $\mathcal{H} := \mathcal{X}_2 \times \mathcal{U}_2 \times \mathbb{R}^n$, and $\mathcal{K}(\zeta) := D\eta(\hat{x}_0, \hat{x}_T)(\xi_0, \xi_T + B_T h)$. Then $\mathcal{K} \in \mathcal{L}(\mathcal{H}, \mathbb{R}^{d_\eta})$ and it is surjective. Analogously, $\mathcal{P}_2^\mu = \{\zeta \in \mathcal{H} : \mathcal{K}^\mu(\zeta) = 0\} = \operatorname{Ker} \mathcal{K}^\mu$, with

$$\|\mathcal{K}^\mu - \mathcal{K}\|_{\mathcal{L}(\mathcal{H}, \mathbb{R}^{d_\eta})} = \mathcal{O}(\mu). \tag{143}$$

Let us now prove the desired stability property. Take $\zeta^\mu \in \mathcal{P}_2^\mu = \operatorname{Ker} \mathcal{K}^\mu$ having $\|\zeta\|_{\mathcal{H}}^\mu = 1$. Hence $\mathcal{K}(\zeta^\mu) = \mathcal{K}^\mu(\zeta^\mu) + (\mathcal{K} - \mathcal{K}^\mu)(\zeta^\mu)$, and by estimation (143),

$$|\mathcal{K}(\zeta^\mu)| = \mathcal{O}(\mu). \tag{144}$$

Observe that, since $\mathcal{H} = \operatorname{Ker} \mathcal{K} \oplus \operatorname{Im} \mathcal{K}^\top$, there exists $\zeta^{\mu,*} \in \mathcal{H}^*$ such that

$$\zeta := \zeta^\mu + \mathcal{K}^\top(\zeta^{\mu,*}) \in \operatorname{Ker} \mathcal{K}. \tag{145}$$

This yields $0 = \mathcal{K}(\zeta) = \mathcal{K}(\zeta^\mu) + \mathcal{K}\mathcal{K}^\top(\zeta^{\mu,*}) = (\mathcal{K} - \mathcal{K}^\mu)(\zeta^\mu) + \mathcal{K}\mathcal{K}^\top(\zeta^{\mu,*})$. Given that $\mathcal{K}$ is onto, the operator $\mathcal{K}\mathcal{K}^\top$ is invertible and thus

$$\zeta^{\mu,*} = -(\mathcal{K}\mathcal{K}^\top)^{-1}(\mathcal{K} - \mathcal{K}^\mu)(\zeta^\mu). \tag{146}$$

The estimation (144) above implies $\|\zeta^{\mu,*}\|_{\mathcal{H}^*} = \mathcal{O}(\mu)$. It follows then from (145) that $\|\zeta^\mu - \zeta\|_{\mathcal{H}} = \mathcal{O}(\mu)$, and therefore, the desired result holds. □

Proof. [of Theorem 9.1] We shall begin by observing that Lemma 9.2 provides a neighborhood $\mathcal{I}$ and a class of solutions $\{(x^\mu, u^\mu, p^\mu, \beta^\mu)\}_{\mu \in \mathcal{I}}$ satisfying (138). We shall prove that $w^\mu = (x^\mu, u^\mu)$ satisfies the sufficient condition (136) close to 0.

Suppose on the contrary that there exists a sequence of parameters $\mu_k \to 0$ and critical directions $(\xi^{\mu_k}, y^{\mu_k}, h^{\mu_k}) \in \mathcal{P}_2^{\mu_k}$ with $\gamma(\xi_0^{\mu_k}, y^{\mu_k}, h^{\mu_k}) = 1$, such that

$$\bar{\Omega}^{\mu_k}(\xi^{\mu_k}, y^{\mu_k}, h^{\mu_k}) \leq o(1). \tag{147}$$

Since $\bar{\Omega}^\mu$ is Lipschitz-continuous in μ, from previous inequality we get

$$\bar{\Omega}(\xi^{\mu_k}, y^{\mu_k}, h^{\mu_k}) \leq o(1). \tag{148}$$

In view of Lemma 9.3, there exists for each k, a direction $(\xi^k, y^k, h^k) \in \mathcal{P}_2$ satisfying

$$|\xi_0^k - \xi_0^{\mu_k}| + \|y^k - y^{\mu_k}\|_2 + |h^k - h^{\mu_k}| = \mathcal{O}(\mu_k). \tag{149}$$

Hence, by inequality (148) and given that $\hat{w}$ satisfies (69),

$$\rho\gamma(\xi_0^k, y^k, h^k) \leq \bar{\Omega}(\xi^k, y^k, h^k) \leq o(1). \tag{150}$$

However, the left hand-side of this last inequality cannot go to 0 since (ξ_0^k, y^k, h^k) is close to $(\xi_0^{\mu_k}, y^{\mu_k}, h^{\mu_k})$ by estimation (149), and the elements of the latter sequence have unit norm. This leads to a contradiction. Hence, the result follows. □

9.2 Control constrained case

In this paragraph we aim to investigate the stability of the shooting algorithm applied to the problem with control bounds (CP) studied in Section 8. Observe that previous Theorem 9.1 guarantees the weak optimality for the perturbed problem when the control constraints are absent. In case we have control constraints, this stability result is applied to the transformed problem (TP) (given by equations (99)-(103) of Section 8) yielding a similar stability property, but for which the nominal point and the perturbed ones are weak optimal for (TP). This means that they are optimal in the class of extremals having the same control structure, and switching times and singular arcs sufficiently close in L_∞. An extremal satisfying optimality in this sense will be called *weak-structural optimal,* and a formal definition would be as follows.

Definition 9.5 (Weak-structural optimality)**.** *A feasible solution $\hat{w}$ for problem (CP) is called a weak-structural solution if its transformed extremal $\hat{W}$ given by* (104)-(106) *is a weak solution of (TP).*

Theorem 9.6 (Sufficient condition for the extended weak minimum in the control constrained case)**.** *Let $\hat{w}$ be a feasible solution for (CP) satisfying Assumptions 8.2 and 8.4. Consider the transformed problem (TP) and the corresponding transformed solution $\hat{W}$ given by* (104)-(106)*. If $\hat{w}$ satisfies* (69) *for (TP), then $\hat{w}$ is an extended weak solution for (CP).*

Proof. It follows from the sufficient condition in Theorem 5.6 applied to (TP). □

Consider the family of perturbed problems given by:

$$\begin{array}{l} \varphi_0^\mu(x_0, x_T) \to \min, \\ \dot{x}_t = \sum_{i=0}^{m} u_{i,t} f_i^\mu(x_t), \quad \text{for } t \in (0,T), \\ \eta^\mu(x_0, x_T) = 0, \\ 0 \le u_t \le 1, \quad \text{a.e on } (0,T). \end{array} \tag{CP_μ}$$

The following stability result follows from Theorem 9.1.

Theorem 9.7 (Stability in the control constrained case)**.** *Assume that the reduced shooting system generated by problem (CP) is square. Let $\hat{w}$ be the solution of (CP) and $\{\hat{T}_k\}_{k=1}^N$ its switching times. Denote by $\hat{W}$ its transformation via equation* (106)*. Suppose that $\hat{W}$ satisfies uniform positivity condition* (69) *for problem (TP). Then there exists a neighborhood $\mathcal{J} \subset \mathbb{R}$ of 0 such that for every parameter $\mu \in \mathcal{J}$ there exists a weak-structural optimal extremal w^μ of (CP^μ) with switching times $\{T_k^\mu\}_{k=1}^N$ satisfying the estimation*

$$\sum_{k=1}^{N} |T_k^\mu - \hat{T}_k| + \sum_{k=1}^{N} \sum_{i \in S_k} \|u_i^\mu - \hat{u}_i\|_{\infty, I_k^\mu \cap \hat{I}_k} + \|x^\mu - \hat{x}\|_\infty = \mathcal{O}(\mu), \tag{151}$$

where $I_k^\mu := (T_{k-1}^\mu, T_k)$. Furthermore, the transformed perturbed solution W^μ verifies uniform positivity (136) *and hence quadratic growth in the weak sense for problem (TP) holds, and the shooting algorithm for* (CP_μ) *is locally quadratically convergent.*

9.3 Additional analysis for the scalar control case

Consider a particular case where the control $\hat{u}$ is scalar. The lemma below shows that the perturbed solutions are Pontryagin extremals for (CP_μ) provided that the following assumption holds.

Assumption 9.8. *(a) The switching function H_u is never zero in the interior of a bang arc. Hence if $\hat{u} = 1$ on (t_1, t_2) then $H_u < 0$ on (t_1, t_2), and if $\hat{u} = -1$ on (t_1, t_2) then $H_u > 0$ on (t_1, t_2).*

(b) If $\hat{T}_k$ is a bang-to-bang switching time then $\dot{H}_u(\hat{T}_k) \neq 0$.

The property (a) is called *strict complementarity for the control constraint.*

Lemma 9.9. *Suppose that $\hat{u}$ satisfies Assumption 9.8. Let w^μ as in Theorem 9.7 above. Then w^μ is a Pontryagin extremal for* (CP_μ).

Proof. We intend to prove that w^μ satisfies the minimum condition (12) given by the Pontryagin Maximum Principle. Observe that on the singular arcs, $H_u^\mu = 0$ since w^μ is the solution associated to a zero of the shooting function. It suffices then to study the stability of the sign of H_u^μ on the bang arcs around a switching time. First suppose that $\hat{u}$ has a bang-to-singular switching at $\hat{T}_k$. Assume, without loss of generality, that $\hat{u} \equiv 1$ on $\hat{I}_k$ and $\hat{u}$ is singular on $[\hat{T}_k, \hat{T}_{k+1}]$. Let us write

$$\ddot{H}_u^\mu = a^\mu + u^\mu b^\mu, \tag{152}$$

where a^μ and $b^\mu := \frac{\partial}{\partial u}\ddot{H}_u^\mu$ are continuous functions on $[0,T]$, and continuously differentiable with respect to μ since they depend on x^μ and p^μ. Assumption 8.5 yields $b^0 < 0$ on $[\hat{T}_k, \hat{T}_{k+1}]$, and therefore

$$b^\mu < 0, \quad \text{on } [T_k^\mu, T_{k+1}^\mu]. \tag{153}$$

Due to (152), the sign of $\ddot{H}_u^\mu$ around T_k^μ depends on $u^\mu(T_k^\mu+) - u^\mu(T_k^\mu-)$. But this quantity is negative since u^μ passes from its upper bound to a singular arc. From the latter assertion and (153) follows

$$\ddot{H}_u^\mu(T_k^\mu-) < 0, \tag{154}$$

and thus H_u^μ is concave at the junction time T_k^μ. Since H_u^μ is null on $[T_k^\mu, T_{k+1}^\mu]$, its concavity implies that it has to be negative before entering this arc. Hence, w^μ respects the minimum condition on the interval $\hat{I}_k$.

Consider now the case when $\hat{u}$ has a bang-to-bang switching at $\hat{T}_k$. Let us begin by showing that $H_u^\mu(T_k^\mu) = 0$. Suppose on the contrary that $H_u^\mu(T_k^\mu) \neq 0$. Then $H^\mu(T_k^\mu+) - H^\mu(T_k^\mu-) \neq 0$, contradicting the continuity condition imposed on H in the shooting system. Hence $H_u^\mu(T_k^\mu) = 0$. On the other hand, since $\dot{H}_u(\hat{T}_k) \neq 0$ by Assumption 9.8, the value $\dot{H}_u^\mu(T_k^\mu)$ has the same sign for small μ. This implies that H_u^μ has the same sign before and after T_k^μ that H_u (before and after $\hat{T}_k$), respectively. The result follows. □

Remark 9.10. *We end this analysis by mentioning that if the transformed solution $\hat{W}$ satisfies the uniform positivity* (69) *for (TP), then $\hat{w}$ verifies the sufficient condition established in Aronna et al. [30] and hence it is actually a Pontryagin minimum. This follows from the fact that in condition* (69) *we are allowed to perturb the switching times, and hence* (69) *is more restrictive (or demanding) that the condition in [30].*

10 Numerical Simulations

Now we aim to check numerically the extended shooting method described above. More precisely, we want to compare the classical $n \times n$ shooting formulation to an extended formulation with the additional conditions on the pre-Hamiltonian continuity. We test three problems with singular arcs: a fishing and a regulator problem and the well-known Goddard problem, which we have already studied in [36, 37]. For each problem, we perform a batch of shootings on a large grid around the solution. We then check the convergence and the solution found, as well as the singular values and condition number of the Jacobian matrix of the shooting function.

10.1 Test problems

10.1.1 Fishing problem

The first example we consider is a fishing problem described in [38]. The state $x_t \in \mathbb{R}$ represents the fish population (halibut), the control $u_t \in \mathbb{R}$ is the fishing activity, and the objective is to maximize the net revenue of fishing over a fixed time interval. The coefficient $(E - c/x)$ takes into account the greater fishing cost for a low fish population. The problem is

$$\left\{ \begin{array}{l} \max \displaystyle\int_0^T (E - c/x_t)\ u_t\, U_{\max} \mathrm{d}t, \\ \dot{x}_t = r\, x_t\ (1 - x_t/k)\ -\ u_t\, U_{\max}, \\ 0 \leq u_t \leq 1, \quad \forall t \in [0, T], \\ x_0 = 70, \quad x_T \text{ free}, \end{array} \right. \tag{P_1}$$

with $T = 10$, $E = 1$, $c = 17.5$, $r = 0.71$, $k = 80.5$ and $U_{\max} = 20$.

Remark 10.1. *The state and control were rescaled by a factor* 10^6 *compared to the original data for a better numerical behavior.*

Remark 10.2. *Since we have an integral cost, we add a state variable to adapt* (P_1) *to the initial-final cost formulation. It is well-known that its corresponding costate variable is constantly equal to 1.*

The pre-Hamiltonian for this problem is

$$H := (c/x - E)\, u\, U_{\max} + p[r\, x\, (1 - x/k) - u\, U_{\max}], \tag{155}$$

and hence the switching function

$$\Phi_t = D_u H_t = U_{\max}(c/x_t - E - p_t), \quad \forall t \in [0, T]. \tag{156}$$

The optimal control follows the bang-bang law

$$\left\{ \begin{array}{ll} u_t^* = 0 & \text{if } \Phi_t > 0, \\ u_t^* = 1 & \text{if } \Phi_t < 0. \end{array} \right. \tag{157}$$

Over a singular arc where $\Phi = 0$, we assume that the relation $\ddot{\Phi} = 0$ gives the expression of the singular control *(t is omitted for clarity)*

$$u^*_{\text{singular}} = \frac{k\ r}{2(c/x - p)U_{max}} \left(\frac{c}{x} - \frac{c}{k} - p + \frac{2px}{k} - \frac{2px^2}{k^2} \right). \tag{158}$$

The solution obtained for (P_1) has the structure **bang-singular-bang**, as shown on Figure 1.

Shooting formulations. Assuming the control structure, the shooting unknowns are the initial costate and the limits of the singular arc,

$$\nu := (p_0, t_1, t_2) \in \mathbb{R}^3.$$

The classical shooting formulation uses the entry conditions on t_1

$$\mathcal{S}_1(\nu) := (p_T, \Phi_{t_1}, \dot{\Phi}_{t_1}).$$

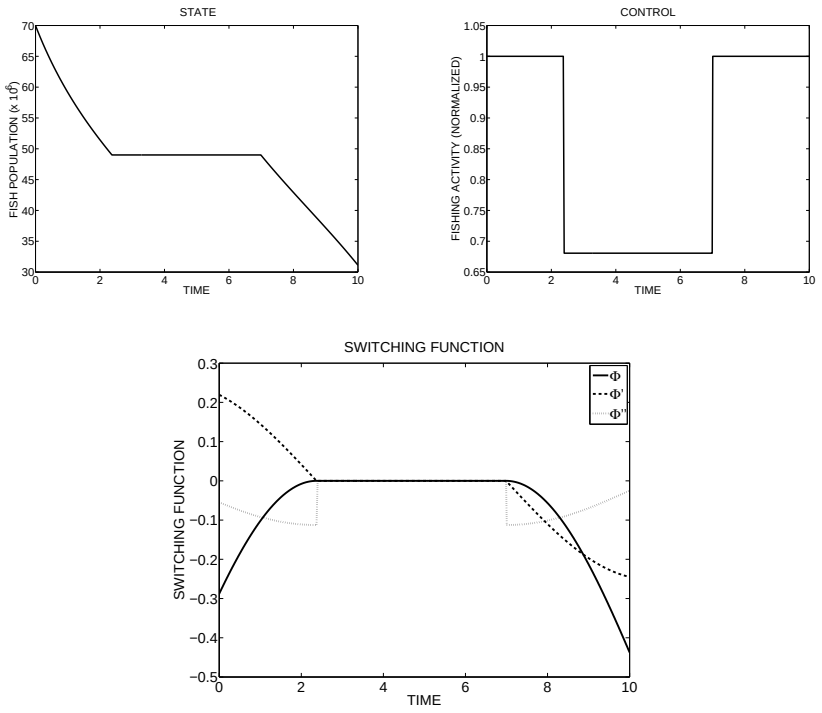

Figure 1: Fishing Problem

Solving $S_1(\nu) = 0$ is a square nonlinear system, for which a quasi-Newton method can be used. Note that even if there is no explicit condition on t_2 in S, the value of p_T does depend on t_2 via the control switch.

The extended shooting formulation adds two conditions corresponding to the continuity of the pre-Hamiltonian at the junctions between bang and singular arcs. We denote $[H]_t := H_{t+} - H_{t-}$ the pre-Hamiltonian jump, and define

$$\tilde{\mathcal{S}}_1(\nu) = (p_{10}, \Phi_{t_1}, \dot{\Phi}_{t_1}, [H]_{t_1}, [H]_{t_2}). \tag{159}$$

To solve $\tilde{S}_1(\nu) = 0$ we use a nonlinear least-square algorithm (see paragraph 10.2 below for more details).

10.1.2 Regulator problem

The second example is the quadratic regulator problem described in Aly [39]. We want to minimize the integral of the sum of the squares of the position and

speed of a mobile over a fixed time interval, the control being the acceleration.

$$\left\{\begin{array}{l} \min \frac{1}{2}\displaystyle\int_0^T \left(x_{1,t}^2 + x_{2,t}^2\right) \mathrm{d}t, \\ \dot{x}_{1,t} = x_{2,t}, \\ \dot{x}_{2,t} = u_t, \\ -1 \leq u_t \leq 1, \quad \text{a.e. on } [0,T], \\ x_0 = (0,1), \quad x_T \text{ free}, \\ T = 5. \end{array}\right. \tag{P$_2$}$$

The corresponding pre-Hamiltonian

$$H := \tfrac{1}{2}(x_1^2 + x_2^2) + p_1 x_2 + p_2 u, \tag{160}$$

and hence we have the switching function

$$\Phi_t := D_u H_t = p_{2,t}. \tag{161}$$

The bang-bang optimal control satisfies

$$u_t^* = -\text{sign } p_{2,t} \quad \text{if } \Phi_t \neq 0. \tag{162}$$

The singular control is again obtained from $\ddot{\Phi} = 0$ and verifies

$$u^*_{\text{singular},t} = x_{1,t}. \tag{163}$$

The solution for this problem has the structure **bang-singular**, as shown on Figure 2.

Shooting formulations. Assuming the control structure, the shooting unknowns are

$$\nu := (p_{1,0}, p_{2,0}, t_1) \in \mathbb{R}^3. \tag{164}$$

For the classical shooting formulation, in order to have a square system, we can for instance combine the two entry conditions on Φ and $\dot{\Phi}$, since we only have one additional unknown which is the entry time t_1. Thus we define

$$\mathcal{S}_2(\nu) := (p_{1,T}, p_{2,T}, \Phi_{t_1}^2 + \dot{\Phi}_{t_2}^2). \tag{165}$$

The extended formulation does not require such a trick, we simply have

$$\tilde{\mathcal{S}}_2(\nu) := (p_{1,T}, p_{2,T}, \Phi_{t_1}, \dot{\Phi}_{t_1}, [H]_{t_1}). \tag{166}$$

10.1.3 Goddard problem

The third example is the well-known Goddard problem, introduced in Goddard [40] and studied for instance in Seywald-Cliff [41]. This problem models the ascent of a rocket through the atmosphere, and we restrict here ourselves to vertical (unidimensional) trajectories. The state variables are the altitude, speed and mass of the rocket during the flight, for a total dimension of 3. The rocket is subject to gravity, thrust and drag forces. The final time is free, and the

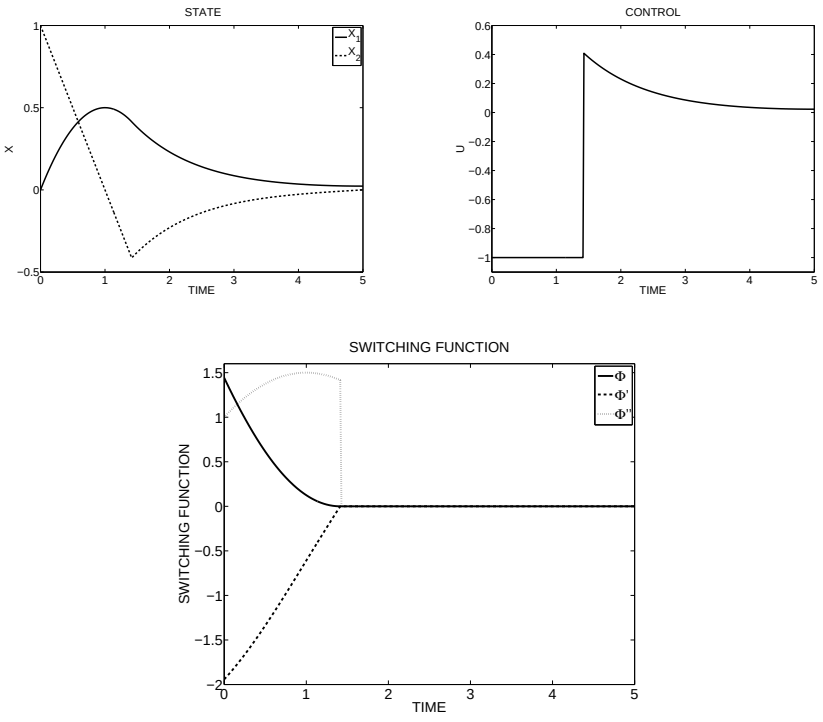

Figure 2: Regulator Problem

objective is to reach a certain altitude with a minimal fuel consumption, i.e. a maximal final mass.

$$\left\{\begin{array}{l} \max\ m_T, \\ \dot r = v, \\ \dot v = -1/r^2 + 1/m(T_{\max}u - D(r,v)) \\ \dot m = -bT_{\max}u, \\ 0 \le u_t \le 1, \quad \text{a.e. on } (0,1), \\ r_0 = 1,\ v_0 = 0,\ m_0 = 1, \\ r_t = 1.01, \\ T \text{ free}, \end{array}\right. \tag{P$_3$}$$

with the parameters $b = 7$, $T_{\max} = 3.5$ and the drag given by

$$D(r,v) := 310v^2 e^{-500(r-1)}.$$

The pre-Hamiltonian function here is

$$H := p_r v + p_v(-1/r^2 + 1/m(T_{\max}u - D(r,v))) - p_m b T_{\max}u, \tag{167}$$

where p_r, p_v and p_m are the costate variables associated to r, v and m, respectively. The switching function is

$$\Phi := D_u H = T_{\max}((1-p_m)b + p_v/m). \tag{168}$$

Hence, the bang-bang optimal control is given by

$$\begin{cases} u_t^* = 0 & \text{if } \Phi_t > 0, \\ u_t^* = 1 & \text{if } \Phi_t < 0, \end{cases} \tag{169}$$

and the singular control can be obtained by formally solving $\ddot{\Phi} = 0$. The expression of u^*_{singular}, however, is quite complicated and is not recalled here. The solution for this problem has the well-known typical structure **1-singular-0**, as shown on Figures 3 and 4.

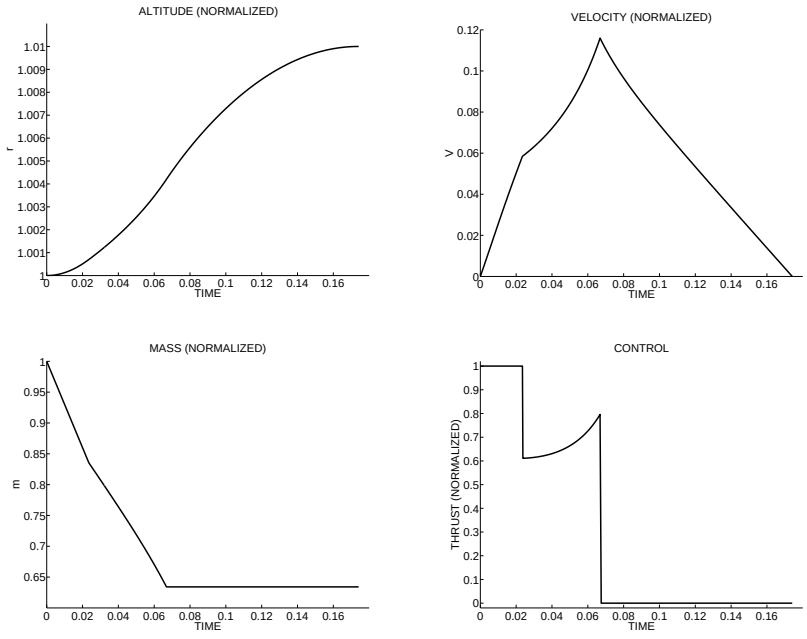

Figure 3: Goddard Problem

Shooting formulations. Once again fixing the control structure, the shooting unknowns are

$$\nu = (p_{1,0}, p_{2,0}, p_{3,0}, t_1, t_2, T) \in \mathbb{R}^6. \tag{170}$$

Here is the classical shooting formulation with the entry conditions on t_1

$$\mathcal{S}_3(\nu) := (x_{1,T} - 1.01, p_{2,T}, p_{3,T} + 1, \Phi_{t_1}, \dot{\Phi}_{t_1}, H_T), \tag{171}$$

while the extended formulation is

$$\tilde{\mathcal{S}}_3(\nu) := (x_{1,T} - 1.01, p_{2,T}, p_{3,T} + 1, \Phi_{t_1}, \dot{\Phi}_{t_1}, H_T, [H]_{t_1}, [H]_{t_2}). \tag{172}$$

10.2 Results

All tests were run on a 12-core platform, with the parallelized (OPENMP) version of the SHOOT ([42]) package. The ODE solver is a fixed step 4th. order Runge Kutta method with 500 steps. The classical shooting is solved

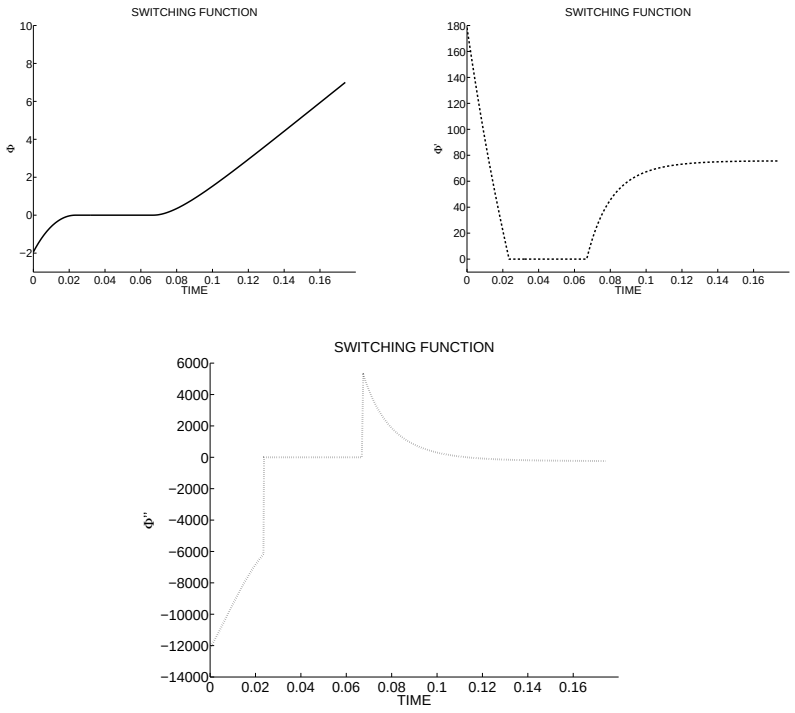

Figure 4: Goddard Problem

with a basic Newton method, and the extended shooting with a basic Gauss-Newton method. Both algorithms use a fixed step length of 1 and a maximum of 1000 iterations. In addition to the singular/bang structure, the value of the control on the bang arcs is also fixed according to the expected solution.

The values for the initial costates are taken in $[-10, 10]$, and the values for the entry/exit times in $[0, T]$ for (P_1) and (P_2). For (P_3), the entry, exit and final times are taken in $[0, 0.2]$. The number of grid points is set around to 10 000 for the three problems. These grids for the starting points are quite large and rough, which explains the low success rate for (P_1) and (P_3). However, the solution was found for all three problems.

For each problem, the results are summarized in 3 tables. The first table indicates the total CPU time for all shootings over the grid, the success rate of convergence to the solution, the norm of the shooting function at the solution, and the objective value. The second table recalls the solution found by both formulations: initial costate and junction times, as well as final time for (P_3). The third table gives the singular values for the Jacobian matrix at the solution, as well as its condition number $\kappa := \sigma_1/\sigma_n$.

We observe that for all three problems (P_1), (P_2) and (P_3), both formulations converge to the same solution, ν^* and the objective being identical to more than 6 digits. The success rate over the grid, total CPU time and norm of the shooting function at the solution are close for both formulations. Concerning the singular values and condition number of the Jacobian matrix, we note that

for (P_2) the extended formulation has the smallest singular value going from 10^{-8} to 1, thus improving the condition number by a factor 10^8. This is caused by the combination of the two entry conditions into a single one that we used in the classical formulation for this problem: as the singular arc lasts until t_f, there is only one additional unknown, the entry time.

Overall, these results validate the extended shooting formulation, which perform at least as well as the classical formulation and has a theoretical foundation.

Remark 10.3. *Several additional tests runs were made using the HYBRD ([43]) and NL2SNO ([19]) solvers for the classical and extended shootings instead of the basic Newton and Gauss-Newton method. The results were similar, apart from a higher success rate for the HYBRD solver compared to NL2SNO.*

Remark 10.4. *We also tested both formulations using the sign of the switching function to determine the control value over the bang arcs, instead of forcing the value. However, this causes a numerical instability at the exit of a singular arc, where the switching function is supposed to be 0 but whose sign determines the control at the beginning of the following bang arc. This instability leads to much more erratic results for both shooting formulations, but with the same general tendencies.*

Problem 1:
Shooting grid: $[-10, 10] \times [0, T]^2$, 21^3 gridpoints, 9261 shootings.

Shooting	CPU	Success	Convergence	Objective
Classical	74 s	21.28 %	1.43E-16	-106.9059979
Extended	86 s	22.52 %	6.51E-16	-106.9059979

Table 1.1: (P_1) CPU times, success rate, convergence and objective

Shooting	p_0	t_1	t_2
Classical	-0.462254744307241	2.37041478456004	6.98877992494185
Extended	-0.462254744307242	2.37041478456004	6.98877992494185

Table 1.2: (P_1) solution ν^* found

Shooting	σ_1	σ_2	σ_3	κ
Classical	3.61	0.43	5.63E-02	64.12
Extended	27.2	1.71	3.53E-01	77.05

Table 1.3: (P_1) singular values and condition number for the Jacobian

Problem 2
Shooting grid: $[-10, 10]^2 \times [0, T]$, 21^3 gridpoints, 9261 shootings.

Shooting	CPU	Success	Convergence	Objective
Classical	468 s	94.14 %	1.17E-16	0.37699193037
Extended	419 s	99.36 %	1.22E-13	0.37699193037

Table 2.1: (P_2) CPU times, success rate, convergence and objective

Shooting	$p_{1,0}$	$p_{2,0}$	t_1
Classical	0.942173346483640	1.44191017584598	1.41376408762863
Extended	0.942173346476773	1.44191017581021	1.41376408762893

Table 2.2: (P_2) solution ν^* found

Shooting	σ_1	σ_2	σ_3	κ
Classical	24.66	5.19	1.96E-08	1.26E+09
Extended	24.70	5.97	1.13	21.86

Table 2.3: (P_2) singular values and condition number for the Jacobian

Problem 3
Shooting grid: $[-10, 10]^3 \times [0, 0.2]^3$, $4^3 \times 5^3$ gridpoints, 8000 shootings.

Shooting	CPU	Success	Convergence	Objective
Classical	42 s	0.82 %	5.27E-13	-0.634130666
Extended	52 s	0.85 %	1.29E-10	-0.634130666

Table 3.1: (P_3) CPU times, success rate, convergence and objective

Shoot.	$p_{r,0}$	$p_{v,0}$	$p_{m,0}$
Class.	-50.9280055899288	-1.94115676279896	-0.693270270795148
Exten.	-50.9280055901093	-1.94115676280611	-0.693270270787320
	t_1	t_2	t_f
Class.	0.02350968417421373	0.06684546924474312	0.174129456729642
Exten.	0.02350968417420884	0.06684546924565564	0.174129456733106

Table 3.2: (P_3) solution ν^* found

Shooting	σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	κ
Classical	6182	9.44	8.13	2.46	0.86	1.09E-03	5.67E+06
Extended	6189	12.30	8.23	2.49	0.86	1.09E-03	5.67E+06

Table 3.3: (P_3) singular values and condition number for the Jacobian

11 Conclusions

Theorems 5.8 and 8.8 provide a theoretical support for an extension of the shooting algorithm for problems with all the control variables entering linearly and having singular arcs. The shooting functions here presented are not the ones usually implemented in numerical methods as we have already pointed out in previous section. They come from systems having more equations than unknowns in the general case, while before in practice only square systems have been used. Anyway, we are not able to prove the injectivity of the derivative of the shooting function when we remove some equations, i.e. we are not able to determine which equations are redundant, and we suspect that it can vary for different problems.

The proposed algorithm was tested in three simple problems, where we compared its performance with the classical shooting method for square systems. The percentages of convergence are similar in both approaches, the singular values and condition number of the Jacobian matrix of the shooting function coincide in two problems, and are better for our formulation in one of the problems. Summarizing, we can observe that the proposed method works as well as the one currently used in practice and has a theoretical foundation.

In the bang-singular-bang case, as in the fishing and Goddard's problems, our formulation coincides with the algorithm proposed by Maurer [5].

Whenever the system can be reduced to a square one, given that the sufficient condition for the non-singularity of the Jacobian of the shooting function coincides with a sufficient condition for optimality, we could established the stability of the optimal local solution under small perturbations of the data.

References

[1] T.R. Goodman and G.N. Lance. The numerical integration of two-point boundary value problems. *Math. Tables Aids Comput.*, 10:82–86, 1956.

[2] D.D. Morrison, J.D. Riley, and J.F. Zancanaro. Multiple shooting method for two-point boundary value problems. *Comm. ACM*, 5:613–614, 1962.

[3] H.B. Keller. *Numerical methods for two-point boundary-value problems.* Blaisdell Publishing Co. Ginn and Co., Waltham, Mass.-Toronto, Ont.-London, 1968.

[4] R. Bulirsch. Die mehrzielmethode zur numerischen lösung von nichtlinearen randwertproblemen und aufgaben der optimalen steuerung. Technical report, Carl-Cranz-Gesellschaft, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany, 1971.

[5] H. Maurer. Numerical solution of singular control problems using multiple shooting techniques. *J. of Optimization theory and applications*, 18(2):235–257, 1976.

[6] H.J. Oberle. Numerische Behandlung singulärer Steuerungen mit der Mehrzielmethode am Beispiel der Klimatisierung von Sonnenhäusern. *PhD thesis. Technische Universität München*, 1977.

[7] H.J. Oberle. Numerical computation of singular control problems with application to optimal heating and cooling by solar energy. *Appl. Math. Optim.*, 5(4):297–314, 1979.

[8] G. Fraser-Andrews. Finding candidate singular optimal controls: a state of the art survey. *J. Optim. Theory Appl.*, 60(2):173–190, 1989.

[9] P. Martinon. Numerical resolution of optimal control problems by a piecewise linear continuation method. *PhD thesis. Institut National Polytechnique de Toulouse*, 2005. http://www.cmap.polytechnique.fr/ martinon/docs/Martinon-Thesis.pdf.

[10] G. Vossen. Switching time optimization for bang-bang and singular controls. *J. Optim. Theory Appl.*, 144(2):409–429, 2010.

[11] M.S. Aronna. Partially affine control problems: second order conditions and a well-posed shooting algorithm. *INRIA Research Rapport Nr. 7764*, October 2011.

[12] B. Bonnard and I. Kupka. Théorie des singularités de l'application entrée/sortie et optimalité des trajectoires singulières dans le problème du temps minimal. *Forum Math.*, 5(2):111–159, 1993.

[13] B. Bonnard, J.B. Caillau, and E. Trélat. Second order optimality conditions in the smooth case and applications in optimal control. *ESAIM Control Optim. Calc. Var.*, 13(2):207–236 (electronic), 2007.

[14] B. Bonnard and M. Chyba. *Singular trajectories and their role in control theory*, volume 40 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2003.

[15] K. Malanowski and H. Maurer. Sensitivity analysis for parametric control problems with control-state constraints. *Computational Optimization and Applications*, 5:253–283, 1996.

[16] J.F. Bonnans and A. Hermant. Revisiting the analysis of optimal control problems with several state constraints. *Control Cybernet.*, 38(4A):1021–1052, 2009.

[17] J.E. Dennis. Nonlinear least-squares. *In: D. Jacobs, Editor, The State of the Art in Numerical Analysis*, pages 269–312, 1977.

[18] R. Fletcher. *Practical methods of optimization. Vol. 1.* John Wiley & Sons Ltd., Chichester, 1980. Unconstrained optimization, A Wiley-Interscience Publication.

[19] J.E. Dennis, D.M. Gay, and R. E. Welsch. An adaptive nonlinear least-squares algorithm. *ACM Trans. Math. Softw.*, 7:348–368, 1981.

[20] A.V. Dmitruk. Quadratic conditions for a weak minimum for singular regimes in optimal control problems. *Soviet Math. Doklady*, 18(2), 1977.

[21] A.V. Dmitruk. Quadratic order conditions for a Pontryagin minimum in an optimal control problem linear in the control. *Math. USSR Izvestiya*, 28:275–303, 1987.

[22] U. Felgenhauer. Structural stability investigation of bang-singular-bang optimal controls. *Journal of Optimization Theory and Applications*, 2011. [published as 'online first'].

[23] U. Felgenhauer. Controllability and stability for problems with bang-singular-bang optimal control. 2011. [submitted].

[24] H.J. Kelley, R.E. Kopp, and H.G. Moyer. Singular extremals. In *Topics in Optimization*, pages 63–101. Academic Press, New York, 1967.

[25] H.J. Kelley. A second variation test for singular extremals. *AIAA Journal*, 2:1380–1382, 1964.

[26] B.S. Goh. Necessary conditions for singular extremals involving multiple control variables. *J. SIAM Control*, 4:716–731, 1966.

[27] J.F. Bonnans. *Optimisation continue*. Dunod, 2006.

[28] E. S. Levitin, A. A. Milyutin, and N. P. Osmolovskiĭ. Higher order conditions for local minima in problems with constraints. *Uspekhi Mat. Nauk*, 33(6(204)):85–148, 272, 1978.

[29] B.S. Goh. The second variation for the singular Bolza problem. *J. SIAM Control*, 4(2):309–325, 1966.

[30] M.S. Aronna, J. F. Bonnans, A. V. Dmitruk, and P.A. Lotito. Quadratic conditions for bang-singular extremals. *Numerical Algebra, Control and Optimization, special issue in honor of Helmut Maurer, and Rapport de Recherche INRIA Number 7764*. [to appear in 2012].

[31] V. Zeidan. Sufficiency criteria via focal points and via coupled points. *SIAM J. Control Optim.*, 30(1):82–98, 1992.

[32] B. Bonnard, J.-B. Caillau, and Emmanuel Trélat. Cotcot: short reference manual. ENSEEIHT-IRIT Technical report RT/APO/05/1, 2005.

[33] A.T. Fuller. Study of an optimum non-linear control system. *J. of Electronics and Control*, 15:63–71, 1963.

[34] J.T. Betts. Survey of numerical methods for trajectory optimization. *AIAA J. of Guidance, Control and Dynamics*, 21:193–207, 1998.

[35] L.T. Biegler. *Nonlinear programming*, volume 10 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. Concepts, algorithms, and applications to chemical processes.

[36] J. Gergaud and P. Martinon. An application of PL continuation methods to singular arcs problems. In A. Seeger, editor, *Recent Advances in Optimization*, volume 563 of *Lectures Notes in Economics and Mathematical Systems*, pages 163–186. Springer-Verlag, 2006.

[37] P. Martinon, J.F. Bonnans, and E. Trélat. Singular arcs in the generalized Goddard's problem. *J. Optim. Theory Appl.*, 139(2):439–461, 2008.

[38] C.W. Clark. *Mathematical Bioeconomics*. John Wiley & Sons, 1976.

[39] G.M. Aly. The computation of optimal singular control. *International Journal of Control*, 28(5):681–688, 1978.

[40] R.H. Goddard. *A Method of Reaching Extreme Altitudes*, volume 71(2) of *Smithsonian Miscellaneous Collections*. Smithsonian institution, City of Washington, 1919.

[41] H. Seywald and E.M. Cliff. Goddard problem in presence of a dynamic pressure limit. *Journal of Guidance, Control, and Dynamics*, 16(4):776–781, 1993.

[42] P. Martinon and J. Gergaud. Shoot2.0: An indirect grid shooting package for optimal control problems, with switching handling and embedded continuation. Technical report, Inria Saclay, 2011. RR-7380.

[43] B.S. Garbow, K.E. Hillstrom, and J.J. More. *User Guide for Minpack-1*. National Argonne Laboratory, Illinois, 1980.

Centre de recherche INRIA Saclay – Île-de-France
Parc Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 Orsay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399